

74. G74/7: P93/2

# THE COMPUTER AND INVASION OF PRIVACY

DOCUMENTS  
ROOM

## HEARINGS

BEFORE A

### SUBCOMMITTEE OF THE COMMITTEE ON

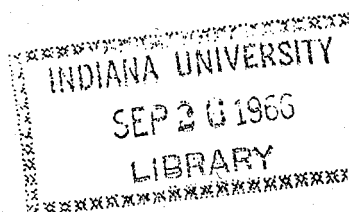
### GOVERNMENT OPERATIONS HOUSE OF REPRESENTATIVES

EIGHTY-NINTH CONGRESS

SECOND SESSION

JULY 26, 27, AND 28, 1966

Printed for the use of the  
Committee on Government Operations



U.S. GOVERNMENT PRINTING OFFICE

WASHINGTON : 1966

## COMMITTEE ON GOVERNMENT OPERATIONS

WILLIAM L. DAWSON, Illinois, *Chairman*

CHEF HOLMFELD, California	FLORENCE P. DWYER, New Jersey
JACK BROOKS, Texas	OGDEN R. REID, New York
L. H. FOUNTAIN, North Carolina	FRANK HORTON, New York
PORTER HARDY, Jr., Virginia	DONALD RUMSFELD, Illinois
JOHN A. BLATNIK, Minnesota	WILLIAM L. DICKINSON, Alabama
ROBERT E. JONES, Alabama	JOHN N. ERLÉNBOEN, Illinois
EDWARD A. GARMATZ, Maryland	HOWARD H. CALLAWAY, Georgia
JOHN E. MOSS, California	JOHN W. WYDLER, New York
DANTE B. PASCELL, Florida	ROBERT DOLE, Kansas
HENRY S. REUSS, Wisconsin	CLARENCE J. BROWN, Jr., Ohio
JOHN S. MONAGAN, Connecticut	JACK EDWARDS, Alabama
TORBERT H. MACDONALD, Massachusetts	
J. EDWARD ROUSH, Indiana	
WILLIAM S. MOORHEAD, Pennsylvania	
CORNELIUS E. GALLAGHER, New Jersey	
WILLIAM J. RANDALL, Missouri	
BENJAMIN S. ROSENTHAL, New York	
JIM WRIGHT, Texas	
FERNAND J. ST GERMAIN, Rhode Island	
DAVID S. KING, Utah	
JOHN G. DOW, New York	
HENRY HELSTOSKI, New Jersey	

CHRISTINE RAY DAVIS, *Staff Director*

JAMES A. LANIGAN, *General Counsel*

MILES Q. ROMNEY, *Associate General Counsel*

J. P. CARLSON, *Minority Counsel*

WILLIAM H. COPENHAVER, *Minority Professional Staff*

## SPECIAL SUBCOMMITTEE ON INVASION OF PRIVACY

CORNELIUS E. GALLAGHER, New Jersey, *Chairman*

BENJAMIN S. ROSENTHAL, New York      FRANK HORTON, New York

NORMAN G. CORNISH, *Chief of Special Inquiry*

CHARLOTTE C. BICKETT, *Clerk*

## CONTENTS

Hearings held on—	Page
Tuesday, July 26, 1966.....	1
Wednesday, July 27, 1966.....	49
Thursday, July 28, 1966.....	119
Statement of—	
Baran, Paul, computer expert with the Rand Corp., Santa Monica, Calif.....	119
Bowman, Raymond T., Assistant Director for Statistical Standards, Bureau of the Budget; accompanied by Paul Krueger, Assistant Chief, Office of Statistical Standards, Bureau of the Budget.....	49
Dunn, Edgar S., Jr., research analyst, Resources for the Future, Inc.....	92
Gallati, Robert R. J., director, New York State Identification and Intelligence System; accompanied by Eliot H. Lombard, special assistant counsel for law enforcement to Governor Rockefeller, and Edward DeFranco, executive assistant to the director.....	146
Horton, Hon. Frank, a Representative in Congress from the State of New York.....	5
Packard, Vance, sociologist, author, and lecturer.....	7
Reich, Charles A., professor, Yale Law School.....	22
Rosenthal, Hon. Benjamin S., a Representative in Congress from the State of New York.....	4
Ruggles, Richard, Department of Economics, Yale University.....	89
Squires, Burton E., Jr., visiting assistant professor of computer science, University of Illinois, Urbana, Ill.....	135
Letters, statements, etc., submitted for the record—	
Gallati, Robert R. J., director, New York State Identification and Intelligence System:	
Act to amend the executive law, in relation to the creation of the New York State Identification and Intelligence System in the executive department and prescribing its powers and duties....	169
Association of the Bar of the City of New York, the.....	181
New York State Identification and Intelligence System.....	159
Report of the Special Committee on Problems of Law Enforcement of the Association of the Bar of the City of New York on proposed legislation to amend the executive law in relation to the creation of the New York State Identification and Intelligence System in the executive department, and prescribing its powers and duties.....	180
Report of the Special Committee on Problems of the Administration of Law Enforcement of the Association of the Bar of the City of New York on the feasibility report and recommendations for a New York State Identification and Intelligence System.....	180
Rundquist, George E., executive director, New York Civil Liberties Union, New York, N.Y., letter to Robert R. J. Gallati, director, Identification and Intelligence Project, dated March 18, 1965.....	158
Sturz, Herbert, the Vera Foundation, Inc., letter to Robert R. J. Gallati, director, Identification and Intelligence Project, dated March 16, 1965.....	159
Supplementary statement.....	179
Horton, Hon. Frank, a Representative in Congress from the State of New York:	
Article from the Saturday Review by John W. Macy, Jr., entitled "Automated Government," dated July 23, 1966.....	35
Excerpt from section 139(b) of title V of the United States Code.....	80

## Letter, statements, etc.—Continued

Michael, Dr. Donald N., professor of psychology and program director in the Center for Research on the Utilization of Scientific Knowledge, University of Michigan, paper entitled "Speculations on the Relation of the Computer to Individual Freedom and the Right to Privacy".....	Page 184
Pemberton, John de J., Jr., executive director, American Civil Liberties Union, New York, N.Y., statement.....	182
Remarks of Hon. Cornelius E. Gallagher and Hon. Frank Horton on floor of the House, August 18, 1966, on questions of invasion of privacy relating to the establishment of a National Data Center....	311

## APPENDIXES

Appendix 1.—Report of the Committee on the Preservation and Use of Economic Data to the Social Science Research Council, April 1965.....	195
Appendix 2.—Statistical Evaluation Report No. 6—Review of Proposal for a National Data Center.....	254
Appendix 3.—The new computerized age.....	295
Appendix 4.—Speech by Vice Adm. H. G. Rickover, U.S. Navy, entitled "Liberty, Science and Law".....	304

## THE COMPUTER AND INVASION OF PRIVACY

TUESDAY, JULY 26, 1966

HOUSE OF REPRESENTATIVES,  
SPECIAL SUBCOMMITTEE ON INVASION OF PRIVACY  
OF THE COMMITTEE ON GOVERNMENT OPERATIONS,  
Washington, D.C.

The subcommittee met, pursuant to call, at 10:10 a.m., in room 2247, Rayburn Office Building, Hon. Cornelius E. Gallagher (chairman of the subcommittee) presiding.

Present: Representatives Cornelius E. Gallagher, Benjamin S. Rosenthal, and Frank Horton.

Also present: Norman G. Cornish, chief of special inquiry; Miles Q. Romney, associate general counsel, Committee on Government Operations; and John Forsyth, special minority consultant.

Mr. GALLAGHER. The subcommittee will come to order.

The Special Subcommittee on Invasion of Privacy today begins its investigation into proposals to establish a National Data Center or Data Bank. The subcommittee is interested in questions of economy and efficiency as they are related to problems of invasion of privacy. The hearings this week continue the special inquiry initiated last year by the House Committee on Government Operations under the authorization of the committee chairman, Hon. William L. Dawson. With me again are the other two members of the special subcommittee, Hon. Benjamin S. Rosenthal and Hon. Frank Horton, both of New York.

Before proceeding with today's testimony, I believe it would be worthwhile to summarize the recently completed first phase of this inquiry by the subcommittee. The investigation centered on three areas: the use of intrusive personality tests by Federal agencies in connection with public employees and job applicants; Federal financial support of research grants and contracts utilizing similar personality tests and questionnaires with children; and a section of the 1964 farm census which asked detailed questions about the outside income and background of all persons living in each farmhouse.

I am happy to report that all of the Federal departments or agencies whose representatives appeared before the subcommittee either have abandoned the use of personality tests, put them on a voluntary basis, or have greatly modified their use.

In addition, Government contractors are now urged to accept the same prohibition on personality tests that the Civil Service Commission applies to Federal employees; and the Bureau of the Census has adopted new procedures to protect the privacy of farmers, members of their families, hired hands and lodgers, when answering questions about outside income in the farm census.



While we are gratified with the successes in these fields, we also are disturbed by other trends within the Federal Government that constitute possible threats to the privacy of American citizens. One such trend is the increasing demand for a centralized facility, within the structure of the National Government, into which would be poured information collected from various Government agencies and from which computers could draw selected facts. It is our contention that if safeguards are not built into such a facility, it could lead to the creation of what I call "The Computerized Man." "The Computerized Man," as I see him, would be stripped of his individuality and privacy. Through the standardization ushered in by technological advance, his status in society would be measured by the computer, and he would lose his personal identity. His life, his talent and his earning capacity would be reduced to a tape with very few alternatives available.

Technology, through the centuries, has enriched human life, and a Federal data center undoubtedly will add to this enrichment and streamline the operation of the National Government. But just as democratic governments historically have secured the freedom of their citizens partly by controlling the fruits of scientific progress, so too must we now make sure that Government computers do not provide the means by which Federal officials can intrude improperly into our lives.

The subcommittee believes it is important that we consider this question before the establishment of a national data center or bank becomes a fact. What we seek at this point is to create a climate of concern, in the hope that guidelines can be set up which will protect the confidentiality of reports and prevent invasion of individual privacy, while at the same time allowing government to function more efficiently and facilitating the necessary research of scholars in statistical analysis.

The problem is potentially serious; its advance solution urgent. It has been reported that funds have already been designated in the 1967 U.S. budget to start a data bank program.

The age of the computer already is upon us. Within the Federal Government alone, it is a billion-dollar business. The Defense Department is devising a computer capable of making 1 billion computations a second. The Federal Reserve Board is gearing plans to make a sharp reduction in check clearance through the use of computers; this in turn will drastically cut the size of the so-called check clearance float. Today's pushbutton telephone will be tomorrow's direct contact with the bank. State and local records are being hooked up on a national scale. Information retrieval in electronic data systems is virtually becoming instantaneous.

Some 20 Federal departments or agencies currently collect and publish data, including the Internal Revenue Service, the Census Bureau, the Office of Education, the Bureau of Labor Statistics, and the Bureau of Old-Age and Survivors Insurance. Information has been supplied to these agencies by persons with the understanding that it will only be used by the receiving agency, for a specific purpose, and in most cases on a confidential basis.

Now, it is suggested that much of this information be pooled in one central source. Presumably, current disclosure restrictions would be adhered to, but the effectiveness of these laws, in some cases, is

dubious. We have even greater fear that the law in some cases is of small concern to those in law enforcement who become overeager in enforcement of that law as they view it. We do not want to see the intended good use of a data center distorted so that it simply makes confidential information more readily available to more people.

Nor do we wish to see a composite picture of an individual recorded in a single informational warehouse, where the touch of a button would assemble all the governmental information about the person since his birth. Such a hypothetical situation could become very real, because into a data bank could be deposited records covering his birth, schooling, military service, employment history, personality traits, credit status, and practically any other aspect of his life. Although the personal data bank apparently has not been proposed as yet, many people view this proposal as a first step toward its creation. I am sure there will be pressures—both within and outside Government—to create one in the future, all, of course, in the interest of economy and efficiency.

The presence of these records in Government files is frightening enough, but the thought of them neatly bundled together into one compact package is appalling. We cannot be certain that such dossiers would always be used by benevolent people for benevolent purposes.

The possible future storage and regrouping of such personal information also strikes at the core of our Judeo-Christian concept of "forgive and forget," because the computer neither forgives nor forgets. We are told that the computer can be programmed to program out derogatory and confidential information; what we fear is the ability to program it in.

We also recognize the obstacles that would be cleared away by a data bank for the researcher, when statistics are instantly available at his asking. But we also recognize the danger implicit in such power which would enable a less scrupulous person—or even a well meaning but overzealous Government official—to delve behind the statistics, to the respondent, and learn the inner secrets of an individual. Shall we create an elite who can narrow and dominate the "corridors of power"? And who shall they be?

We also are aware that vast governmental projects are being undertaken to help develop America, and that the Federal Government seeks new information about social and economic conditions to plan and operate these programs. But such programs should not be at the cost of individual privacy.

What we are looking for is a sense of balance. We do not want to deprive ourselves of the rewards of science; we simply want to make sure that human dignity and civil liberties remain intact. We would like to know just what information would be stored in a national data center; who would have access to it; who would control the computers; and, most importantly, how confidentiality and individual privacy would be protected. Thought should be given to these questions now, before we awaken some morning in the future and find that the dossier bank is an established fact, and that liberty as we knew it vanished overnight.

Perhaps new guidelines and safeguards will have to be established to correspond with the sudden development of the computer. Perhaps a new concept in file accessibility will have to be formulated, under

which a person may have the right to examine his own file, or at least have the right of appeal, so that he can insure the accuracy and completeness of his dossier.

These are some of the questions to which we are seeking answers. We believe that once both sides—the need for technological advance and the right of privacy—have been presented before this subcommittee, their inseparability will become obvious, and a sense of balance hopefully will be achieved.

The issue is not whether a statistical data bank can be established nor whether it would be beneficial. A statistical data bank can be established and great benefits can be derived from it.

However, there appears to be a great imbalance between technology on the one hand, and the law and public interest on the other. The issue is, therefore, can we achieve a balance so as to assure that technological progress will serve man and that man's free will will dominate in the new environment that the computer is rapidly bringing about?

Mr. Rosenthal, you have a statement?

#### STATEMENT OF HON. BENJAMIN S. ROSENTHAL, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

MR. ROSENTHAL. Mr. Chairman, at the outset of our second series of hearings, I would like to reaffirm my appreciation of your efforts, and Chairman Dawson's wisdom in establishing this special inquiry. I believe we have already had considerable impact—and creative influence which brings honor to the entire Committee on Government Operations, and the Congress as a whole.

My own personal reaction to the proposal for a National Data Center was, I suppose, similar to that of most citizens—intense apprehension at the prospect of still more invasions of personal privacy. In so many areas technological progress is being secured at the expense of personal liberty. The projected National Data Center seems an almost too fitting symbol for that development. And yet, I admit, there is the continual danger of excessive reaction and inordinate fear. For the problem of balancing the benefits and costs of progress is subtle and requires careful study and prudent judgment.

I approach the issue with initial skepticism. I have yet to be convinced of the necessity for a central bank of highly personal data on all American citizens. I have yet to learn why each agency cannot maintain its own files. And I find it hard to believe that the improved efficiency afforded by the Center would outweigh the clear risks.

Even with the most precise safeguards, we must continue to ask ourselves certain basic questions. Is the increased threat to personal liberty too great a price to pay for the anticipated efficiency and progress? Are we sacrificing too many aspects of our personal lives for limited objectives? Does the additional knowledge we might gain yield benefits to society greater than the losses to the individual?

I intend to ask these questions of all witnesses before this committee. I will not be satisfied with any witnesses who favor the establishment of a National Data Center and who fail to answer such questions satisfactorily. I propose to inquire if adequate safeguards can be formulated so that we can benefit from the growth of technology in

the area of personnel data control. But their development is an absolute precondition to the establishment of any projects such as that before us today. I think I speak for the chairman and for my colleagues in stating that we will tolerate no unnecessary intrusions into the privacy of American citizens, regardless of their source and nature.

Mr. GALLAGHER. Mr. Horton.

#### STATEMENT OF HON. FRANK HORTON, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

MR. HORTON. Mr. Chairman, the mission of this subcommittee, investigating instances of individual privacy invasion caused by or contributed to as the result of Federal Government action, is important and timely. Clearly, our experiences and endeavors of the past year, have proved this point. And I feel your exposition of the subcommittee's work sets forth with special significance the wisdom of Chairman Dawson in chartering this subcommittee.

Privacy, as a fundamental freedom of the American citizen, is an unquestioned constitutional right. That this subcommittee, through examination and exposure, has curbed a brand of overzealousness on the part of certain Government agencies to overlook this right in personality testing is a notable example of the inherent protections to be found in our Federal system of checks and balances.

As significant as those earlier hearings were, I have become convinced that the magnitude of the problem we now confront is akin to the changes wrought in our national life with the dawning of the nuclear age. Proposals to gather in one central location or in one giant data bank all the information which Federal agencies amass on the citizens of this country are sufficiently filled with possibilities for privacy invasion that I believe it is eminently proper for our subcommittee to conduct this investigation.

These data bank concepts are a product of modern technology. Today the computer is a central figure in our society. The increasing rate at which it will change our lives exceeds the imagination, exceeds even the imagination of the computer men who foster it. Dr. Jerome B. Wiesner, dean of science at MIT and former science adviser to President Kennedy, has said.

The computer, with its promise of a millionfold increase in man's capacity to handle information, will undoubtedly have the most far-reaching social consequences of any contemporary technical development. The potential for good in the computer, and the danger inherent in its misuse, exceed our ability to imagine \* \* \*. We have actually entered a new era of evolutionary history, one in which rapid change is a dominant consequence. Our only hope is to understand the forces at work and to take advantage of the knowledge we find to guide the evolutionary process.

We will be fortunate if we are able to keep these processes "evolutionary" and not "revolutionary."

Assuming the best for a moment, let us regard our computer systems as good and fair and the computer men behind the console as honest and capable. Even in these circumstances, there is danger that computers, because they are machines, will treat us as machines. They can supply the facts and, in effect, direct us from birth to death. They can "pigeonhole" us as their tapes decree, selecting, within a narrow range, the schooling we get, the jobs we work at, the money we can earn and even the girl we marry.

It is not enough to say "It can't happen here"; our grandfathers said that about television.

Now, let us compound the concern. Assuming a computerman who was dishonest, unscrupulous or bent on injury, there would be nothing sacred. We could be destroyed.

Adm. Hyman Rickover has expressed a fundamental concept concerning these problems: he states that we must realize that the power of these computers is technology, and technology must serve man; man must never blindly accept technology, he must take up the challenge and control it. It is a force he has to master and use to his benefit.

The admiral exhorts us to be faithful to individual basic values, to preserve our right of privacy and independence and to bend this fantastic new technology to our principles. It is the function of lawgivers, in Admiral Rickover's view, to set the limits within which computermen can operate. He makes it clear that this is not a limit on science or knowledge but only on our use of knowledge and technology.

The concept of such control is ancient. Fire controlled is our friend; uncontrolled it is devastating. The wheel is man's servant and yet his greatest exterminator. The computer is another two-edged sword. It will take more than the controls of the "horse-and-buggy" days to use computers for our benefit and yet keep them from making shreds of human dignity, privacy, and freedom.

To provide an example, despite the flood of technical language some Government consultants use to camouflage their recommendations, the fact remains that a central data service bank would require:

One, that confidential information now in Government files would be forwarded to a new group and used for other purposes than it was originally given; and

Two, that a new group would have the code and would know the names, addresses and background of the people who submitted the confidential information.

Tying the two together would be an easy matter.

It is held that personal dossiers are not intended, but no thoughtful computerman can deny that they are a logical extension of present plans. I am pleased to say that computermen as a group are deeply concerned with the problem of controlling information storage and retrieval so that no one ever will be able to take away our basic freedoms through these means.

One last point: The argument is made that a central data bank would use only the type of information that now exists and since no new principle is involved, existing types of safeguards will be adequate. This is fallacious. Good computermen know that one of the most practical of our present safeguards of privacy is the fragmented nature of present information. It is scattered in little bits and pieces across the geography and years of our life. Retrieval is impractical and often impossible. A central data bank removes completely this safeguard.

I have every confidence that ways will be found for all of us to benefit from the great advances of the computermen, but those benefits must never be purchased at the price of our freedom to live as individuals with private lives.

Mr. GALLAGHER. Thank you, Mr. Horton.

The subcommittee is very fortunate this morning in having as its first witness in this series of hearings Mr. Vance Packard. Mr. Packard is more responsible than any man in our country for alerting us to the dangers that lurk in the twilight of our sophisticated society and the changes that are coming about and creating perhaps a new environment.

Vance Packard, for the last decade, has served as a public conscience for various actions by this Government and business that infringe on individual rights and individuality, largely through such books as "The Hidden Persuaders," "The Status Seekers," "The Wastemakers" and "The Naked Society."

We are indeed honored and privileged to have you open these hearings this morning, Mr. Packard.

Please proceed.

#### STATEMENT OF VANCE PACKARD, SOCIOLOGIST, AUTHOR, AND LECTURER

Mr. PACKARD. Mr. Chairman and members of the committee, thank you for inviting me to submit my thoughts on the matter of your current concern, the proposals to establish within the Federal Government national data centers.

The chairman has already identified me. I would only say my particular concern has been the impact on the individual citizen of pressures generated by social and technological change, and that my last book, "The Naked Society," was specifically concerned with the erosion of individual privacy arising from these changes and I have talked and written on the basis of later developments which, as you well know, have frequently involved activities within the Federal Government.

I believe your inquiry is an immensely important one. The implications of permitting the Federal Government to assemble in central data banks, or a center, information provided by or about its individual citizens are far reaching. There is clearly a threat to individual liberty in such a project.

Some time ago Dr. Robert Morison, as scientific director of the Rockefeller Foundation, warned:

We are coming to recognize that organized knowledge puts an immense amount of power in the hands of the people who take the trouble to master it.

Certainly we are seeing and hearing that information is power. If this is so, then all of us should be uneasy about the vast amount of information that the Federal Government is starting to store away on its citizens in dossiers, card files, and electronic memory banks.

In recent years we have seen an enormous growth in techniques for observing, assessing, controlling, keeping track of and exchanging information about people. To mention just a few, we have the hidden tape recorders, the lie detectors, the undercover agents, the personality assessors, and the bureaucratic investigators. Many of the techniques in use have been made much easier for the trained investigators by the proliferation of recordkeeping in our increasingly bureaucratic society.

Recordkeeping itself is becoming fantastically more efficient in reducing space and promoting speedy retrieval of information.

have moved from the use of bulky folders devoted to individuals to coded file cards, thence to punch cards that can be machine sorted, then to the tape used in electronic memory banks. And now the experts are learning that it is feasible to hook up a number of electronic memory banks, data banks, to one giant system which can be activated for purposes of inserting or retrieving information from numerous locations.

Records on individual Americans now number in the billions. We leave a trail of records behind us from the moment of birth. Our birth, incidentally, is recorded not only on a birth certificate, but also on our parent's income tax return when we became a deductible item.

Consider for a moment this trail of records most of us leave. Non-Government filekeepers have, for example, the information on our income, worth of home, debts, and location of our bank which we often surrender when we apply for credit. Then there are the reports made to insurance companies investigating us as risks. This may involve an appraisal of our social life and our sex life as well as our financial stability. There are the employment files that may be complete with results of personality inventories and lie detector tests. There are hospital and medical records, and the records of the moving companies that have prudently made an inventory of our possessions being moved.

Units of our State and local governments have our school records including our grades, IQ, and any reports of emotional difficulties. At least one of them will have our driving record, any brushes with the law, our property holdings and all licenses obtained, including any for marriage or its dissolution.

The Federal Government has our tax returns over a number of years, our responses to the increasingly lengthy census questionnaires, our social security record, our application for a passport, and perhaps our fingerprints. If we have been in military service, worked for a defense contractor or for the Federal Government, there are lengthy files on us that may well indicate known associates, affiliations, religious beliefs. If we have applied for an FHA loan on a home there will usually be an estimate of the prospects that our marriage will hold together.

These are just a few of the records we leave behind. Many of the filekeepers have been quick to learn that the records they control may hold considerable interest for people outside their own organization. Years after our birth, for example, an interested party such as a lawyer may be happy to pay \$50 for information from our birth certificate, which officially is confidential information. And in a number of cities there are entrepreneurs happy to obtain and sell the information, as well as hospital records, police records, immigration records, passport records, and so on.

Then there are the legitimate organizations engaged in selling personal information about us. The credit bureaus circulate to members the information that we give a store when we seek credit; and if there is any interest the same information may be passed on to curious Federal investigative agencies that inquire. Also there are the giant investigating enterprises such as the Retail Credit Co. which has openly assured prospective clients that it can strengthen any report it makes by drawing from its vast file of investigations made

in previous investigations. The investigative firms making reports for insurance companies on insurance applicants have often felt free to sell the same information to noninsurance clients.

Within the Federal Government agencies have increasingly been developing systematic patterns for exchanging information. When a Federal agent makes a national agency check on a person, it customarily involves checking the files of at least eight Federal agencies.

The individual citizen who is concerned about the erosion of his privacy has up until now had some consolation in the knowledge that all these files about his life have been widely dispersed and often difficult to get at. Digging up a sizable file of any individual has been a time-consuming, expensive proposition.

This is changing with the advent of giant computers with their capacity for instant recall of a great variety of available information. The Federal Government has led the way in installing larger and more sophisticated computers. It has purchased many thousands of computers, including some of the world's largest, and the Civil Service Commission is now operating a training center which has taught 2,300 Federal employees how to get maximum usable information out of various computer systems. The Internal Revenue Service's massive investment in computers to store and assess information on taxpayers is well advertised. We can all be cheered by the promised increase in fairness and efficiency in tax collecting that presumably will result, but if we look forward into the future a decade, the prospect is disquieting. For each taxpayer there presumably can be developed in the electronic memory banks a "cum" or cumulative file covering up to 10 years of his life. Not only would a vast amount of information he has provided about himself and his family over the decade be subject to virtually instant retrieval, but also, theoretically, information about known associates during the period and people and organizations who have had business dealings with him. In short, there will be the capacity, at least, for an instant dredging of one's dimly remembered personal affairs of the past. Unless procedures are developed to prevent unreasonable harassment through this capacity, then the taxpayer of 1976 may well be in a poor mood to celebrate the bicentennial of the Declaration of Independence.

Which brings us to the proposal to consolidate some of the major Federal filing systems into one vast central data bank "by using the new information technology now available." Consultants have urged this concept of centralizing data upon the Bureau of the Budget. A special task force has now been appointed to study the idea. At first this central storage center would pool information now in the files of 20 different Federal agencies. The agencies were not specified in the announcement creating the task force but the pooling reportedly will include records from the Internal Revenue Service, the Census Bureau, the Bureau of Old-Age and Survivors Insurance, and the Federal Reserve Board, to mention just four. Presumably in future years more and more agencies—and more and more of their records—would enter the pool.

The announcement implies the Government is interested only in assembling statistics in more readily available form. But every one of us in this room is a statistic, especially if the statistic involving us has our social security number attached. One consultant to the Bureau of



the Budget has been quoted in the Washington Post as urging that valuable information is lost if confidences are kept and statistics are made anonymous too early in the game. No secrets would be kept from the central data center. The raw data about people's lives would be fed into the central computers without concern for confidentiality and the computers would be programmed to act as the censors. If the Government is sincere in saying it is interested only in generalized statistics, then it would seem essential that all individual identification of a statistic be removed before the kernel of desired information is fed to any central computer system.

But apparently more than one central data center is envisioned, and at least one would deal with live people, not depersonalized statistics. In the Saturday Review of this past week John W. Macy, Jr., Chairman of the U.S. Civil Service Commission, details in glowing terms the wonders of "Automated Government." In talking of the Federal Government's monumental job in keeping track of all its employees, he proposes that—

We must have integrated information systems. This will require the use of information across departmental boundaries \* \* \* Direct tape-to-tape feeding of data from one department to another may become common.

He and others talk of the great gains to come from centralizing data about millions of people in or out of government. There would be the broadening of the horizons of knowledge, the greater efficiency, the dollars saved.

We should be wary of promises that the goals of such consolidation of data are only modest ones that would interest statisticians and planners. Unless there are safeguards, pressures will surely grow to assemble more and more specific data about specific individuals. When the social security program began we were assured that our social security number would be guarded as a secret so that no one could possibly use it to keep track of our movements. Today we must write our social security number not only on our income tax return, but must supply it to banks holding our money and to organizations making payments to us. Our social security number in fact is so easily obtainable that one nationwide investigating firm has a line on its standard form where the investigator must list the social security number of the person he has investigated.

Or consider the census. The authors of the U.S. Constitution called for an "enumeration" of the population every 10 years. But by 1960 the census has gone far beyond enumeration. Many millions of citizens in 1960 had to answer 165 questions about their lives, purchasing habits and incomes. And the pressure is growing to add a host of new inquiries such as ethnic origins, religious affiliation, schooling, et cetera, to the 1970 census. Failure to answer every question the census director decides to ask you can result in a fine or jail sentence.

We should also be concerned about what seems to be a lack of sensitivity among some administrative officials about the implications to the individual involved of becoming computerized by the Federal Government. The announcement of the creation of the task force detailed several points to be studied but no mention was made of exploring the impact on the citizen. And Mr. Macy in his enthusiastic descriptions of automating the Federal processing of personnel, said that the Government must ask: "What parts of the job can a com-

puter do better—and which parts can men do better." He is presumably influenced by the per-dollar performance criterion so popular now in Washington.

In all these plans for centralizing data about citizens it seems to me that the crucial question is whether we are letting technology get out of hand without being sufficiently concerned about human values.

There would seem to be a number of hazards in the growing fascination of the Government's electronic file keepers with the idea of exchanging and pooling data on the lives of our citizens.

First, I believe the concept of having a central data bank for use in making decisions involving citizens threatens to encourage a depersonalization of the American way of life. Our people increasingly, and rightly, are resenting their treatment as numbers being controlled by a computer. This resentment is believed to be a factor in the student unrest at the vast State universities, where the student's exams may be machine-graded and his ID number is often printed twice as large as his name.

As for the management of Government personnel by one super-machine, it should be noted that even the best of machines cannot assess the unique personality, the special talents, the particular aspirations and motivations of the individual citizen. And all the recent evidence indicates that any normal citizen performs best where he is encouraged to take personal responsibility for handling a special task in his own unique way.

Mr. Macy is pleased to report that in our automated Government hundreds of thousands of people already are being largely hired by machine. The applicant makes marks on paper in a form that can be digested by the computer. The machine grades him, decides what the passing grade will be for any particular batch of applicants, and then writes a letter advising each applicant if he has passed or failed. I am reminded of a description given me of some of the new automated food canneries. A truck brings baskets of peaches up to the automated building. Once the peaches are unloaded automated machinery takes over. The peaches are washed, peeled, sliced, pitted, siruped, canned, packed in cartons entirely by machine. The cartons, properly labeled and consigned, emerge out the other end of the building.

Most of us applaud the automated processing of peaches. But does it follow that we should applaud the automated processing of people? I think not.

A second obvious hazard inherent in the central data banks is that they will increase the distrust of the citizens in their own Government and alienate them from it. People are becoming wary of what they tell their Government as they discover that information they are confiding for one purpose may be used to affect their life in some entirely different connection. If what they tell the FHA to get a home loan turns up as a knockout factor when they seek a job with a Government contractor, they will start being wary. And they will warn their friends to be wary.

In addition there will inevitably be a suffocating sense of surveillance as the public learns that their Government is developing an all-seeing eye. In the past, one of the hallmarks of totalitarianism, whatever its particular form, has been this sense that somewhere there is an all-seeing eye.

A third hazard inherent in the central data banks is that they will greatly increase the likelihood that the life chances of many citizens will be unfairly affected. A central file can absorb large batches of data about people but it is ill equipped to correct errors, allow for extenuating circumstances, or bring facts up to date. An acquaintance related to me his wrath when he discovered, quite by accident, that his local credit bureau, in a litigation report on him, listed him as having been the target of three lawsuits for failure to meet commitments. On the record he obviously was a person to beware.

The facts were that the first was a \$5 scare suit back in the 1930's over a magazine subscription he had never ordered. Nothing came of it. The second involved a disagreement over a \$200 lawyer's fee. It was later compromised amicably and withdrawn. The third involved a disagreement with a client over a fee he had charged, for services rendered. This was the only one of the three that actually got to court and he had won it. It took the man 2 days of digging to clear his record by proving to the bureau's filekeepers the disposition of the three suits.

Many employers including the Federal Government require a job applicant to note if he has ever been held by a law-enforcement agency for investigation. In recent years hundreds of thousands of citizens have been held momentarily for investigation and then released with no charges made. It is hard to explain to a computer feeder the innocent circumstances.

Even more serious in affecting one's life chances is the fact that the computer is incapable of making allowances for early errors or indiscretions. It has no capacity to recognize that people indeed often do change and become more responsible as they grow up. The son of a friend in a Midwestern city applied at several department stores in the area for a job when he was graduated from high school at the age of 18. He had recommendations from his minister, Scoutmaster, high school principal, and chief of police. But no store would even give his application serious consideration. It turned out that his name was in a central file maintained for the stores, possibly a computerized one, of known lawbreakers. Five years earlier, at the age of 13, the boy, while still figuratively in short pants, had been caught snitching \$2 worth of fishline from a store.

America's frontiers were largely settled by people seeking to make a fresh start. They were often seeking to get away from something unpleasant in their past, either painful episodes, misdemeanors, poverty, or oppression. Today with episodes of our past increasingly being recorded in central files and computers the possibility of the fresh start is becoming increasingly difficult. The Christian notion of the possibility of redemption is incomprehensible to the computer.

Finally there is the hazard of permitting so much power to rest in the hands of the people in a position to push computer buttons. When the details of our lives are fed into the central computer where they are instantly retrievable, we all to some extent fall under the control of the machine's managers. Public figures running for office in opposition to allies of the machine's managers possibly could be smeared with information from the computer at a point where there is inadequate time to set the record straight. In recent years we have seen at least one notable case in Washington where information from

a secret dossier was used to try to intimidate or discredit a person who was making statements embarrassing to certain high officials. The filekeepers of Washington have derogatory information of one sort or another on literally millions of citizens. The more such files are fed into central files, the greater the hazard the information will become enormously tempting to use as a form of control.

Some highly knowledgeable people have been issuing warnings about the hazards residing in exploitation of the computer. These are the computer designers themselves. One computer maker, Bernard S. Benson, has conceded that the concentration of power in the form of accumulated information can be "catastrophically dangerous." And at a meeting of computer experts sponsored by New York University, another expert, computer scientist Eldridge Adams, said that the electronic computer systems being used by Government agencies were collecting so much information about families and individuals that those who controlled the machines were achieving a "truly frightening" power. He said that without proper control the computers could convert our society into the Big Brother regime predicted in the book "1984."

Let us remember, 1984, is only 18 years away. My own hunch is that Big Brother, if he ever comes to these United States, may turn out to be not a greedy power seeker, but rather a relentless bureaucrat obsessed with efficiency. And he, more than the simple power seeker, could lead us to that ultimate of horrors, a humanity in chains of plastic tape.

I would hope that this committee, by its power to influence governmental procedures and to propose legislative safeguards—and the Congress, by its power to legislate and to exercise oversight—will take action promptly to assure that such a state can never come to pass in this land.

Thank you very much.

Mr. GALLAGHER. Thank you very much, Mr. Packard, for your statement which gets to the heart of the fears that many of us have as we view the escalation of efficiency and economy that is always possible and probable in looking down the road of computer progress.

The real problem is not that a computerized data bank can be created. I think all of us agree that it can. There are some reasons why it should.

What this committee is seeking is suggestions you might offer as to how safeguards—as to what safeguards—can be installed in the computer process.

I am not satisfied in my own mind that such safeguards can be properly built in, but if they can be, what would you suggest?

Mr. PACKARD. If they can be installed, certainly I would suggest as a minimum that they must be. I think that the Congress should in its functioning as a watchdog of the executive arm of the Government, make sure that information does not leave one department to go into a central file until individual identification is removed. How this would be involved in Mr. Macy's proposal to make one vast Federal file of Government employees, I do not know. The only heartening thing about his proposal is that it does not involve all American citizens; it simply involves the millions who are involved in the Federal Government.

But even in his case, I think it is urgent that this committee look into the proposition very carefully and look into the possibilities of safeguards in terms of information being passed from one department to another on what he calls a tape-to-tape basis, across departmental lines.

Mr. GALLAGHER. Is it your feeling then that any central data bank should confine itself to statistics that are sanitized of individual identification?

Mr. PACKARD. That is certainly correct, sir, yes.

As I understand the technical problems, this would be a little more expensive, but certainly if the Government wants to get into assembling statistics that can be used by the Government's planners in terms of our larger scale programs, I think that individual identification must be removed and I don't think the Government should be involved in pooling information where such information cannot be removed—where the identification cannot be removed.

Mr. GALLAGHER. Of course, this is one of the problems, as we view it—the hearings are just beginning—the more safeguards that you attempt to build in, the higher the cost becomes.

Therefore, what we fear is a quick and dirty computer—quick and dirty in the line that it would be a system quickly put together and not free of the safeguards, or rather, not having the safeguards that are so necessary to protect the citizen.

It is my understanding in the computers we use for defense, the most expensive part of it becomes the safeguards that are necessary to be installed, both from a hardware standpoint and a software standpoint, and that when you get into this area you increase the cost from 3 to 10 times the cost of the system itself, but herein lies the danger.

What would your comment be as far as the central data bank or the credit collection system or the files that now rest in the Government are concerned, as to a law that might allow an individual citizen to have access to that information which pertains to him? Therefore, he could check it for accuracy and this would keep the system a little more honest.

Mr. PACKARD. I was greatly impressed by this suggestion when you first made it. I think it is an excellent proposal. I think also that if it is inaugurated, the individual citizen should be provided information on how he can make the check without going through a maddening runaround in Government bureaus to try to get information.

I have to give my social security number to many different organizations now and I am a little obscure on what my number really is, although I have been using it for many years, so I wrote to social security asking for a confirmation. This was more than a month ago and I still haven't gotten any word. I don't quite know how to go about it. I simply wrote to Baltimore and that is as far as I know.

Mr. GALLAGHER. Mr. Packard, you cite some hair-raising facts and statistics in your book, "The Naked Society," concerning the compilation of statistics and dossiers on millions of Americans in the Federal Government. Your citations include the exchange of information by Government bodies that possess confidential records and the swapping of information between Government file builders.

This committee has had considerable examples that have been cited where there was a free and easy exchange of information—lie detector evaluation, psychological test results—between personnel officers.

Have you any more recent information in this area and could you explain what implications they may have in connection with the proposed centralization of these dossiers in the Federal Government?

Mr. PACKARD. I have a great deal more recent information that does not spring readily to mind in organized form. Certainly it is true, as you indicate, that the results of lie detector tests have been passed around, and I know at some of the credit bureaus there are regular desks where Federal agents can come and take information from the credit bureau forms.

You also have, of course, the exchange of information on criminal activities among about eight different Federal agencies involved in having any information about crime. I suppose we should applaud this where you are dealing with known criminals, although I think they also exchange information about suspected criminals. Perhaps this is defensible as one of the areas in this computerizing and centralizing of information.

In general, when you are dealing with the citizen as a free citizen, I think he is entitled to know the information he gives to one arm of the Government is not going to be used against him by another arm of the Government when he assumes the information was given confidentially.

Mr. GALLAGHER. You have mentioned the inherent fear of our citizen. This may well be a cause for the unrest that prevails in our country today. The specter of a human being coping with a fellow human being is not insurmountable, but now that he is overawed because he must now cope with a machine plus a human being, perhaps this awe is justified, but is this justification, in your opinion, something that we must accept, or is it possible for the man of the future to cope with a machine which will control his destiny?

Mr. PACKARD. I think that is the fundamental question, Mr. Chairman. I do not think the answer must necessarily be a negative one. I think Aldous Huxley before he died said that just because there are blind forces at work on the lives of individuals, this does not necessarily mean the individual cannot protect himself. I think all of us as Americans and as individuals—and this includes members of the Government and Members of the Congress and citizens and individuals—need to be aware of the fact that we do have the blind force of the computerized age that we must cope with, and we must put boundaries around it so it will serve us usefully rather than become a tyrant.

Mr. GALLAGHER. Dr. Westin, an expert on the problems of privacy, says:

Unless the issue of privacy is in the forefront of the planning and administration of future computer systems, the possibilities of data surveillance over the individual in 1984 could be chilling.

Would you care to comment on this?

Mr. PACKARD. I certainly would agree that is correct. Your earlier comment about the awe of computers I think is important, because people are more frightened of things they cannot understand, and I cannot understand the computer myself. I have been through the IBM factory and have had it explained to me and I have talked to

many computer experts, but I would not have any assurance that I know what it is all about, it is changing so rapidly and so swiftly. I think we do have this threat, as Dr. Westin said and as many others have said, hanging over us that the machine can dominate our lives.

Mr. GALLAGHER. There is a certain mystique that the computer experts have created that all will be well if we just leave it in their hands and that we are creating a "we" and "they" society, "we" being the statistics and "they" being those who will see that all is well.

One of the things that this committee is attempting to ferret out is whether or not there is a mystique and whether or not we should trust and rely on this mystique or whether or not it is time that we start taking a hard look at just what the position of the law is, what the position of the citizen is, in relation to what is now a very definite possibility of a central data bank, that central data bank being the step in the door toward the possibility of a personal dossier on all of our citizens in the record collection you have outlined.

Mr. PACKARD. I think it would be ideal if the committee can have the guidance of computer experts who are not associated with organizations that have a vested interest in selling computers to the Government or, in fact, selling computers anywhere. Those experts are at the universities, and I think they could be very helpful.

Mr. GALLAGHER. Mr. Horton.

Mr. HORTON. Thank you, Mr. Chairman.

Mr. Packard, I think there are two things that stand out in your testimony. The first is the detailing of the information that is already available to the Federal Government and other agencies about private citizens. Then the second aspect is that even without the computer there has been a trend away from the confidentiality of information the Federal Government has obtained about citizens.

You spoke with regard to the social security number and the magnification of the loss of private rights there. I think these two points you underscore in your statement are very helpful to the committee.

I think also you have pointed out in your statement that up to this date, at least, there has been a dispersal of this information which in and of itself has tended to give some protection to the individual.

I think one of the big problems we have in this whole area is to comprehend what these computers can do. Would you like to make any comment with regard to that, based on your experience and your study?

Mr. PACKARD. Thank you, Mr. Horton. As I understand it, you will be having further experts on computers who can speak more authoritatively on the capacity of the computer as of July 1966. The capacity of the computers is changing not just by years but by months. I would hesitate to make any comment beyond the fact of what the computer experts themselves are saying, that we are progressing not only from getting information on individuals, but building giant systems—to me this is the fascinating or the frightening thing—giant systems where information can be either put in or retrieved from a number of different locations, including distant ones and even including telephoning information into the computer or calling and getting information by telephone out of the computer. This, I think, is probably the most frightening thing of all in terms of getting the control of the computer out of hand and getting more and more people having

access to the information so security does obviously become more difficult to maintain. If you have a lot of people who can get into the computer and take out information, it always will be harder for any investigating body later to try to trace down how the leak occurred.

Mr. HORTON. I was not so much interested in having you comment on the technology of the computers because I realize that is something the experts will have to do. I was thinking more in terms of your opinion as one who has delved to a great extent into this problem of the invasion of the right of privacy, your comment on how you as a layman, not an expert in the technology of computers, envision what will be happening to the individual citizen with the use of this computer system.

For example, consider how many 12-digit numbers can be added in 20 seconds. With a pencil, the answer is two. With an adding machine, 10. With a computer, 160 million. The rate of change is astronomical. This change is on us already. It is here.

I was more interested in your layman's view as to what is going to happen to the individual with the use of this type of highly technical information or tool which can compile this information and have it available. I was more interested in your personal opinion, rather than the technical aspects of it.

Mr. PACKARD. Certainly there are many functions in our society that can be greatly improved by the use of computers. As a matter of fact, I am using a computer myself in terms of some research I am getting together to work out correlations and findings. So, I do not think we should be frightened by the computer as a machine, but I think we should be frightened by what it can do if the information involves individuals and their identity, and I think we need to make sure that before the Federal Government gets too deeply involved in installing computers that have data fed into them involving individuals' records, that this body and the Congress make sure that the individual identification is removed from the material.

Mr. HORTON. You would not, certainly, curtail the population information about people?

Mr. PACKARD. No; certainly not.

Mr. HORTON. You are not advocating that computers not be used.

Mr. PACKARD. Oh, certainly not; no.

Mr. HORTON. You are advocating that this technological advance be harnessed and that there be established safeguards to protect the individual.

Mr. PACKARD. That is correct.

Mr. HORTON. One of the proposals or suggestions which you would make would be along the lines of that which the chairman has made, to give some sort of right to the individual to know what information is available on him so he can correct it or have his own opportunity to be heard, as it were.

Mr. PACKARD. That is correct, yes.

Mr. HORTON. Have you any other thoughts with regard to how this advance could be harnessed and safeguards built into it?

Mr. PACKARD. No thoughts occur to me offhand, no.

Mr. HORTON. Have you any thoughts or any possible recommendations with regard to how the Congress could act in this field?

Mr. PACKARD. I think the Congress and this committee and several of the congressional committees have been extremely effective in their



roles of alerting, simply by making Federal agency directors aware of the human implications of what they are doing, as, for example, the success which has been achieved in persuading the various departments to make less use of lie detectors and less use of personality testing, the post office to stop using the mail cover, the Justice Department to stop using wiretapping, and many agencies to stop using the snooper buttons. All these have come about not by legislation but by the mere fact that a body such as yours has explored the situation and made the administrators in the various Federal agencies aware, perhaps for the first time, that there are human value issues involved, and this is what they have been doing, and that they have usually responded by establishing safeguards.

Mr. HORTON. Thus, you feel the action of committees having hearings such as this focusing attention on the problem will have some benefit?

Mr. PACKARD. I certainly do, yes, sir.

Mr. HORTON. You referred to Mr. Macy's comments in an article in the Saturday Review. What was your reaction to this? Do you feel the Government is going too far in this automated process in regard to civil service?

Mr. PACKARD. I think Mr. Macy is a fine man but—

Mr. HORTON. I am not talking about him personally.

Mr. PACKARD. I think he is overly fascinated with the wonders of electronics in terms of reducing the cost of sorting personnel and processing personnel. Since the Federal Government is involved in dealing with millions of people, he sees it in terms of millions, and dollar costs per unit. I think from the standpoint of all that is being learned by the behavioral scientists on human motivation and what makes people perform better, people perform better if they are given a chance to perform in a unique way, and also they perform best if they have a sense that they are being trusted. When people have a feeling that they are not being trusted, they tend indeed to become more untrustworthy. This is what I think you had. The agencies that were using the lie detectors and similar things were simply digging the holes deeper because they were making people more untrustworthy by that fact.

Mr. HORTON. Another point you indicated earlier was the removal, if possible, of the individual identification. It seems to me this would be difficult to accomplish, or else the information would not be beneficial to the Federal Government. It seems it is an impractical possibility.

Mr. PACKARD. As I understand it, the Bureau of the Budget has proposed the central data bank for reasons of overall planning, rather than for information about individuals. Ostensibly, there would be no gain to the Bureau of the Census in having the names attached to all this information it is getting. The problem is apparently it would be difficult to wash out this information from the tapes if it were fed into a central computer. That is the heart of the problem, I believe.

Mr. HORTON. That is all the questions I have, Mr. Chairman.

Mr. GALLAGHER. Thank you, Mr. Horton.

I agree that is the heart of the problem, and that is what we are trying to spotlight today—the necessity of eliminating the individual name if we are interested merely in statistics for problems of planning.

If this were not one of the safeguards and names and social security numbers remained on the tapes, do you see a real threat here, or are we getting prematurely overheated in our concern?

Mr. PACKARD. I do not think you are getting prematurely overheated at all. I think we should all be scared stiff about the possibility that these giant machines would be fed data about individual Americans and that this information would be retrievable by a number of different organizations or groups. I think this would clearly create the preconditions for a totalitarian system.

Mr. GALLAGHER. Looking down the path, assuming we get beyond 1984, is it not one of the great responsibilities of our Government to guarantee and protect human values, and is it not necessary at this point to start programing our own Government toward this end?

Mr. PACKARD. I certainly agree, yes. I think any government that has control over a people has at the minimum a responsibility to treat all individuals involved with a sense of decency and dignity. I do not think you would have this sense of decency and dignity if we are treated as numbers in a memory bank and under the control of a giant computer center.

Mr. CORNISH. Mr. Chairman, I have only one question in the interest of time.

Mr. PACKARD. I was very much interested in the example you gave in your testimony of your friend and his trouble with the credit bureau. I think in answer to one of the chairman's questions you mentioned that in some credit bureaus they actually have separate desks where Federal investigative agents can come in and sit down and make whatever notations they want to.

Mr. PACKARD. Yes.

Mr. CORNISH. We would assume, then, whatever notations they made or whatever copies they made would become, presumably, a part of a Federal file somewhere.

Mr. PACKARD. That is correct, yes.

Mr. CORNISH. If we examine the idea of a personal dossier bank, then it would be a possibility at some time in the distant future that the very information which was obtained from a private credit bureau could end up in a Federal dossier center.

Mr. PACKARD. Very definitely, yes, sir.

Mr. CORNISH. That leads me to ask you this question. Do you think this might be the proper time for the Congress to go back to the original source of the information in the credit bureau and possibly institute some new safeguards affecting credit bureaus? What I am suggesting is perhaps a requirement that an individual be permitted to examine his own credit report on file with the credit bureau to determine its accuracy and completeness.

Mr. PACKARD. I think that would be an excellent idea. I think also it would be very helpful if the Congress did look into the practices of these investigating agencies such as the credit bureaus that accumulate fantastic amounts of data about individual citizens and, in the absence of stiff regulations controlling them, have rather relaxed policies about exchanging information with other people.

Mr. CORNISH. You have devoted some discussion in "The Naked Society" to the credit bureaus and their reports. Would you agree that credit bureau reports are definitely in the stream of interstate commerce?

Mr. PACKARD. That is a very provocative thought, and certainly true because the credit bureaus as a matter of function exchange information in a network all across the country. More than 100 million records are tied in through the exchange of information that is available.

For example, if a man from the State of Nebraska moves to the State of Connecticut and applies to a store in Connecticut for credit, the credit bureau can put in a request for the Nebraska credit rating on him. So this is an interstate operation.

Mr. CORNISH. In your experience, this information does cross State lines?

Mr. PACKARD. Very definitely. Millions of items a year cross State lines; yes, sir.

Mr. CORNISH. Thank you very much.

Mr. GALLAGHER. Mr. Romney.

Mr. ROMNEY. Just one question, Mr. Packard. I think we can see from your statement that you would have reservations about a central personnel data center. Do you actually oppose the concept within the Federal Government of a central personnel data center for Government employees?

Mr. PACKARD. Yes. I think in the announcement that Mr. Macy made, in his enthusiasm he did not put enough emphasis upon the hazards involved in terms of human values and the spread of information. He seemed to assume that every bit of information that any branch of the Government gets on an individual should be pooled. I would have reservations about that, yes, and would view it with considerable concern. I would hope that the committee would do that, too.

Mr. ROMNEY. That is all, Mr. Chairman.

Mr. GALLAGHER. Mr. Forsyth.

Mr. FORSYTH. I am a little confused by some of the mixing of terms and agencies and everything else. It seems as though we are mixing sensitive and insensitive information together; if there is such a distinction. We are also mixing the type of people and agencies that can withdraw information. We are also, I think, mixing up—at least, it seems to me to be confused a bit—whether or not the names are linked to the information.

As I understand it, to go back down the order in reverse, there is no way that a central data system can be effective for a multitude of purposes unless the names are linked to the information at some point in some computer, even though perhaps not in the active set of files or active tape going through the machine; is that correct?

Mr. PACKARD. I do not know. I think if the Bureau of the Budget is sincere in what it says is its aim in developing this central data system, it is not interested in individuals but is interested only in Government broad-scale statistics that can be used.

Mr. FORSYTH. They say that is true, but they say in order to build a system which is responsive to the needs of users being interrogated by a number of different agencies with different questions, the computer has to go back to the basic building block, the individual, in each case, and even though it does not have to identify him by name, it has to identify him by social security number or something so they do not get duplication of material.

For instance, when new material comes in, it has to be related to that particular individual and not his neighbor, and it cannot be re-

ported on twice. So, each time the computer has to verify back the individual and ask a new set of questions. So, even though the operators themselves may not know the name, even though the information relating to that person will never come out under his name but only as a statistic, the system still has to know his name. I believe that is correct, is it not?

Mr. PACKARD. I do not think the whole system would have to know his name. I think the agency contributing the information would need to know his name.

Mr. FORSYTH. By "system," I meant the one main computer.

Mr. PACKARD. No; I do not think so. I think you could have an automatic altering of numbers and setting up a new system of numbers so they would not be tied to the social security number.

For purposes of consolidation, you could assign any individual a number that would not be revealing of anything about him except this would go back to the same building block.

Mr. FORSYTH. Within that computer. Somewhere in the central system, some group of people have to have a code which codes that number to the person.

Mr. PACKARD. It should be the department or agency that supplied the information in the first place. I think you do get the hazard when there is a central knowledge of the individuals.

Mr. FORSYTH. The problem with that is that another agency sends in information, some of which is duplicative and some not. It is fed into a central computer, and if it is set up under a new number, then you have the same number reporting twice, with some duplication. So the same number has to come each time from computer to computer.

Mr. PACKARD. That I think would be the heart of the problem. I do not know the technicalities of it. I think this would be a good point to explore with computer experts rather than with me.

Mr. HORTON. I think one of the points you are trying to make is if there is any way to disperse the information within the computer bank, this at least would be some type of safeguard, just as the dispersal of this information now without computers is somewhat a safeguard to the individual.

Mr. PACKARD. That is correct; yes, sir.

Mr. HORTON. You do not know technically how this can be accomplished.

One other thing I wanted to ask you about. Do you feel there is a basis for the Congress perhaps, or whoever is in control of this system, limiting or having a limit placed on it as to what data can be placed in the bank? Do you see what I mean?

Mr. PACKARD. Yes, I think this should be a concern of this committee. What type of data can be properly and safely put into central data systems should be a concern of this committee.

My own feeling is that you are getting into danger as soon as you get any data that can be identified in terms of an individual citizen or taxpayer.

Mr. HORTON. You made the point with reference to the census, that it was basically, in the beginning, just an enumeration of the population, but now it has gotten into a lot of personal questions. The thought I was trying to follow with you with respect to this is whether

or not we should limit the data and information concerning the individual and use that technique as perhaps a means of safeguarding and controlling.

Mr. PACKARD. I think you are correct. I think you would be safer if you are going to start a filing system that is going to lead to a central filing system, the identifications should be removed at the original agency before the material goes on tape.

Mr. HORTON. Thank you.

Mr. GALLAGHER. Mr. Packard, you hoped the committee would look into Mr. Macy's suggestion that there be computerization and central files on Federal employees. I assure you if such a system is put together, it is our hope and the insistence of this committee that each employee will have access to his own file to see what is in there so that it is accurate and honest. I think this is the only way that an employee would be able to cope with all the information-gathering services so he would be put on notice and would have an opportunity to examine what has been collected on him.

Mr. PACKARD. I think this committee would be performing a very great service if it could persuade the Civil Service Commission to make that safeguard available.

Mr. GALLAGHER. I want to thank you very much for your appearance here this morning, Mr. Packard, for your alerting our country in the past long before many people got to thinking about these problems, for your contribution to the people of this country, and for honoring us here with your presence this morning. Looking down the path, if people read all of your books—I hope they do—and we put this question up to a national referendum, I think there would be some predictable results. We want to thank you very much.

Mr. PACKARD. Thank you, sir.

#### STATEMENT OF CHARLES A. REICH, PROFESSOR, YALE LAW SCHOOL

Mr. GALLAGHER. The Chair would like to call Prof. Charles A. Reich. In behalf of the committee, we welcome you here this morning, Professor Reich.

Professor Reich is from the Yale Law School. He is an expert on the collection of data on individuals and an expert on the legal implications of the collection of dossiers. We have asked Professor Reich to relate his experience of the past to the new problem of the computer and the possibility of computerized dossiers on citizens and taxpayers in the United States.

Professor Reich, we welcome you; would you please proceed?

Professor REICH. Thank you very much, Mr. Chairman.

I might say that the field I teach in is constitutional law. I will talk today about some of the legal aspects of the proposals that have been made.

I might also say I am here in my own behalf, not on behalf of the law school or Yale University or anybody except myself.

When I began to think about this problem, it occurred to me that everybody is in favor of privacy. I noted that within recent weeks, Time magazine, Saturday Review, Newsweek, and all of my friends said they favored privacy, and I believe that you would be able to get a 100-percent vote out of the American people on the same subject.

The trouble comes in particular instances. The trouble is particular invasions of privacy all too frequently take place because, although we believe in it in general, we are frequently unable to see why a particular invasion should be prohibited. Every time a proposal comes along that would invade privacy in one way or another, there is somebody to say it is necessary, there is somebody to say that the information obtained is very reliable and important, and we have to have it, and there is somebody to say that there is no really important new invasion of rights, that is to say, there is someone who will explain that this does not make a very great new intrusion into the life of the individual.

Thus it is that devices like the lie detector, wiretapping, and eavesdropping are all too frequently a matter of practice because we think or somebody thinks that we have to have the information that can be obtained.

These particular devices have been discussed so much that I thought in talking about privacy today I would like to talk about the proposed Federal data center and use that as the way of talking about this problem of necessity, this problem of whether we have to have this particular invasion, too.

I do not know exactly what is proposed in a Federal data center. I will just have to guess. I will just have to assume that the various kinds of information that the Government now has in many different places and many different bureaus relating to school, relating to employment, relating to the Army, to criminal convictions, recommendations that come in from many sources about a person—that all of that would be centralized somewhere and available to authorized agencies or persons. I do not know who they would be.

Mr. GALLAGHER. If I could narrow the issue, Professor, what is before this committee now for our consideration is a proposal by the executive branch for the establishment of a central data bank that would centralize the information collected by 20 agencies of the Government which now collect and publish information, such as the Census Bureau, Social Security, and many of the other agencies. Many of these are now protected by law as to confidentiality.

Along with the proposal is the warranty, for whatever it is worth, that the law will be adhered to and that there will be a protection of the individual who has given the information.

What is now before this committee and what it is concerned with is that while there are now 20 agencies that collect and publish information, there are many agencies of the Federal Government which collect and do not publish information, and we view this as a foot-in-the-door proposition through which eventually the collectors, but not particularly the publishers of information, will also use it in the interest of efficiency and economy. I think what we are considering here is a new proposition, certainly something new in the Federal Government, a departure from the original right of protection of the individual. That is the issue before the committee.

Professor REICH. Many thanks. I take it we are talking primarily about information that does exist somewhere now, and a proposal to centralize it.

One of the things we have to deal with is the argument that the individual be no worse off after this proposal than before because all the

information exists now. I think that is where I feel like taking off on my own opinion.

Mr. GALLAGHER. We hope you will, because that is what we are here to listen to.

Professor REICH. The first thing which seems to me basic about dossiers and the centralization of information is that information gets less reliable the further away it is from the source. I always use this as an axiom, a proposition that I would say is true all the time. Let me give you a college illustration to show what I mean.

When our freshmen come in the college, we have upper-classmen as counselors for them, and the counselors are supposed to show them around and tell them where the laundry is and where to go to find Vassar College, and so forth. It is generally a very helpful thing. Part of their duties as counselors is to write a report on each freshman, which goes in a file. They will say whatever comes into their heads about the freshman, that he is nice, that he is not nice, that he has a chip on his shoulder, whatever they feel is an accurate report. That is something between them and the dean's office at Yale College.

If this information gets into a file and stays there, it slowly changes from a reasonably accurate statement to something that could be a tremendous falsehood, because the freshman counselor disappears and we do not know any more who he is if we want to find him to ask him what he meant. The freshman may change. Perhaps he had a chip on his shoulder the first year, but after he got to like the place he got to be one of the nicest guys around. So, the information may no longer be true except as of the time that it was made. Other people reading this may understand it in other ways.

If that particular record, which serves ordinarily a good purpose, were to survive for 10 or 15 years, were it to become a part of other people's information, it seems to me it would become an untruth by the passage of time and by the distance from the source. Indeed, those who have looked at freshman records many years later are often shocked by the idea that they still exist, and wonder why they have not been destroyed.

That is my basic proposition about the development of inaccuracy, and it is my answer to the idea that this data center would be reliable. I do not think it would be reliable.

Freshman records are only something that happen between freshmen and their counselors. I would like to tell you a little about what happens between me and my students, because this is directly related to the data center. I get—and I am not unusual in this respect—all kinds of questionnaires about my students. I am supposed to fill them out and send them often to private people, but much more often to some Government agency. I take it this is the raw material, or some of it, out of which the computer center would have its supply.

For instance, here is the Department of Health, Education, and Welfare asking me about Student "X" whom I know reasonably well. It asks this kind of question: "How do you rate the applicant's relationships with other people? Consider such things as ability to work and get along with superiors and subordinates." Then it gives you answers. For example, you can check this one: "Mediocre. Wants to do things his own way more often than is desirable. Disliked by some associates. Somewhat lacking in tact. Becomes sullen when criticized.

Tends to react negatively to suggestions." There are other proposed answers all the way from "outstanding" to "poor."

Another question: "How do you rate the applicant's personal adjustment? Consider such things as emotional stability and maturity." Let's again take the mediocre suggested answer: "Inclined to be nervous, irritable, easily annoyed. Inclined to get feelings hurt. Has little insight regarding personal limitations."

My point about this is the same one I made about the freshman. That is, first of all, do I really know what I am talking about when I check these things off? Maybe this fellow was in the day before and was in a bad humor, had a fight with his girl friend, and so he was easily annoyed, nervous and irritable. Maybe the next day he would be a charmer all the way through. Maybe I am mad at him for some reason. Maybe I had a headache the day I checked this thing off. At any rate, this information petrifies a momentary flash reaction on my part. I am assuming I am not malicious, assuming I am a fairminded person trying to do a good job. However fair I try to be, this information goes to the Government and there it is, and no power on earth can change it one iota.

It seems to me if I were a psychiatrist, maybe I would be qualified to answer questions like this, but as a professor of law I do not really know anything about the personal adjustment of my students. I do not really know anything at all about their relationships to other people. I see them in a very special situation in the class, and I am asked to say all kinds of things, good or bad, about subjects that I know nothing whatever about.

If I do not fill these out, somebody somewhere in Washington is going to say, "There must be something to hide. This professor is unwilling to answer these questions."

Just think of the poor student who maybe talked only once in class in the whole year and said something that made him look very nervous. He is a nervous man forever, although that may be the one night in the year that he did not get a good night's sleep, or something else.

So, I am talking about the kind of things that ask for more than we know and then make it into the truth.

The Peace Corps is always after me. The Peace Corps asks for a rating on emotional maturity. Remember, again, I am a teacher. I do not know, I would suppose, very much about emotional maturity, but I am allowed to check things like "Candidate is emotionally unstable, has a history of emotional outbursts, withdrawal, other signs of inability to cope with stress."

I am asked, again, for judgments far outside my ability. I must check something. I have no choice. The student may not know anyone else in school as well as he knows me. I am on the faculty. Whatever I say is there forever.

Again, a Peace Corps rating on relationships with other people. I am allowed to say "Superb." How could I say "Superb"? I do not know how he treats fellow students outside of my presence. "Poor." "Doubtful." "Reasonable to believe candidate will have difficulty working with others." Maybe he has trouble working with me. All I can say is that I must answer, and whatever I say is there.

Mr. HORTON. May I ask the witness to establish the basis on which he gets these questionnaires? Are they sent to you in connection with



job applications, or is this something that comes to you on all students who are in your class?

Professor REICH. Mr. Horton, every student who graduates from law school has to get a job someplace, and nearly every place requires references. In the case of the Government, he must give them the names of those teachers he knows best. So, on all the students I know, I receive from private or public agencies questions which are at the student's behest but, nevertheless, he is required to give references. He probably chooses me because he knows me and he hopes I will say something nice about him. Sometimes a student comes in and says, "You don't remember me, but you gave me an 'A' and I would like a recommendation." I say, "Did I? Well, you must be a good student."

You see, I do not even know all the people that well, but the system requires that I appear to. It is in connection with employment that this crosses my desk.

Mr. GALLAGHER. I am sympathetic with the problem, Professor, because we get the same questions as Congressmen, and the only related incident and the whole reason we are asked to serve as references is the fact that a mother or father might have voted for us. We have to fill out the whole business concerning character.

Please continue.

Professor REICH. I want to illustrate this one more time by the Navy, which gives you a series of boxes to check. As to each thing you can say "outstanding," "excellent," "good," "satisfactory," "unsatisfactory." For example, "Ability to lead others," "Degree of cooperation with others," "Emotional characteristics and stability," "Attitude toward carrying out desires of those in authority over him."

Look at the danger of that question, for example. I am in authority over these students. Maybe I am awfully unreasonable. Maybe once in a while they say, "I don't think this ought to be done the way you say. I have a better idea how to write this paper." Is that something that is going to get my back up and am I going to check him "poor" on this?

The point is, if somebody wants to come and talk to me about a student, inside of 10 minutes they get an idea of what I am like and they get an idea of the way in which I talk about people. But a machine does not know those things. So when I say "unsatisfactory" because the man talked back to me, the machine only knows that this man is unsatisfactory. It does not know anything about me. Maybe it is I that is unsatisfactory and not the student. All of that is part of this whole problem. There is no way to go back to the source.

Another source of information that gets into the files is something I would call private adjudications, that is, formal decisions about people that are made outside of the courts. For example, we, the Yale Law School, find that a student cheated and that becomes part of his permanent record. The problem there is to know whether we did as good a job as we should have to find out if it was true that he cheated. We pride ourselves, because we are a law school, on being very careful. We hear the student. We give him every opportunity to explain. We try to find the facts as lawyers should. But schools all over the United States engage in the process of disciplining students, and not all of them are lawyers and not all of them know how to do things fairly. Some of them do not hear the student at all.

Again, what validity do these private decisions have? They can be a curse on the individual for the rest of his life, but you may not have any idea whether they are really accurate or not. They may meet no standards of fairness with which we are familiar.

So, in each case of information other than the formal decisions of courts, we do not know what is really meant by the information in the files. As it gets step by step farther away in distance and further away in time from the original source, it becomes less and less accurate until what was the truth can become a lie.

About people in Government, about workers in Government—I do not want to use the word "bureaucrats" because it is a bad word—I would just like to say when anybody who works in a great organization receives information from a computer center and it says "unsatisfactory" in this category, "unreliable" in that category, it takes a lot of courage for a Government employee to say, "I am going to hire him anyway. I am going to disregard this. I think he looks good to me. I have seen him. I don't care what this professor said in New Haven."

Most people in Government, because it is so big, because they are part of a great chain of responsibility, are going to say, "I don't want to take a chance." That is the common reaction of a man in a big organization. So when they see something bad, they are going to say, "I don't trust my own judgment. I had better trust what is on paper, because if we hire this young student, whom we like, and he does something wrong, he turns out to be a loser after a while, it will be my fault. I saw this in the record and I failed to stop it. I am going to be blamed. The safe thing is just to say 'No,' and hire a man with an unblemished record."

So, every normal human reaction is going to be to give more weight to these things in the file than I as the maker of the file ever meant. Often I might check something off and I would like to say to the man who is going to hire this fellow, "Disregard this. We didn't get along, the two of us. I would give him a chance if I were you." But I never get to say that.

So the reaction of the normal person who reads the file is to say, "I don't want to take a chance."

I have examples, also, of the inquiries that come at the other end. They do not happen to be from Government because I have no access to Government inquiries, but these are inquiries from private people to the Yale file. For example, let us imagine that Yale has its own computer center. It does not. It just has a file. Somebody writes in to the dean and says, "The person whose name appears above has applied to us for a position. We would like any information regarding his scholastic standing, character, and personal habits. Your reply will be kept in strictest confidence." This comes from businesses, and so forth. This is what I imagine would be the inquiry to the computer center.

Notice that they want anything of a pertinent nature about the personal habits, and so forth. When it comes to Yale, what happens? Well, the dean is not going to answer all these letters. They have a student often, maybe a third-year law student or somebody like that, who does part-time work in that bureau. All he does is go and look at the files of the applicant and pull out anything that he thinks these people

would like to know. So, there is no more control, even at a good university, over this sort of thing than just the judgment of a person in a clerical role who can go and look at all these data and send it forth.

Likewise, it seems to me, the tendency would be for a sort of automatic dispersal of information without any judgment being made as to its quality, because even at Yale we do not judge its quality. If we send it out we just say, "This is in the file, and we suppose this is what you would like to know."

So it is that what one puts in comes out, but it does not come out the same as what was put in.

That is the situation, as I see it, of the data center. The question is: How does this stand with respect to law? How does it stand with respect to the Constitution?

At the very beginning of my thoughts about the law, it seems clear to me that any time bad information is supplied about an individual, his legal rights are invaded at that moment. We have a great common law tradition that defamation of character is a wrong in the law. It seems to me absolutely clear as a starting proposition that anybody who supplies derogatory information about somebody else invades his legal rights. He may have no remedy. That is something that happens in the law. But the beginning of my thinking is that a person has a right not to be defamed, whether it is by a machine or by a person.

The second thing that seems to me so crucial here is that this whole process is secret. The individual does not know what I have said about him. He does not know what is in the computer's file. He does not know what the computer says about him. He does not know what judgments people make on the basis of that.

I think this is a denial of the constitutional right to confront, the constitutional right to face those who make statements about you, to question them, and to rebut, to answer. It is elementary, it seems to me, that this right is lost in the kind of case that I have given.

Suppose for some unfortunate reason I have a grudge against a student and, receiving one of these things, I check "unsatisfactory," and so on; the story, if it came out, would be a very different story if he were able to cross-examine me. He could show there was some bad blood between us or something else, and I was not being wholly accurate. The truth, as lawyers know, is brought out in an atmosphere of adversary proceedings, of cross-examination, of being able to answer, to rebut.

Here we have what seems to me over and over again instances of condemnation without trial, of information supplied without confrontation, and of a denial to the individual of any chance whatever to answer.

The Supreme Court has recently been extraordinarily scrupulous with respect to the right to have a lawyer and the right to confront in situations where people are charged with crime. The right to have a lawyer and the right to confront in situations where people are charged with crime. The right to have a lawyer, for example, begins now at the very earliest moment of contact in the police station.

Here are people who are not even charged with crime, and yet who may be punished far more severely than the ordinary criminal. Here are people whose opportunity to have jobs, to earn money, whose reputations and everything else are about to be damaged forever, and they

have no trial, no lawyer, no opportunity to find out anything. It seems to me without question a denial of due process of law to send forth bad information about a person in secret in that way.

It is in this that I see the essence of the evil of the automatic data center. It is in this notion of the petrification; that is, this man is called bad by somebody, hence he is bad forever, and there is nothing he can do about it. There is no remedy in the law.

It seems to me that we deal, when we think about a computer center, with a wrong which the law is unable to make right in any way that we now know. It is a form of damage without a remedy, and it can be a very terrible punishment, indeed, for the individual.

Beyond the invasion of the right not to be defamed is a second, and some people think, a more vague right, the right to privacy itself.

When information about a person of the type I have mentioned is distributed all over to everybody, certainly you could say at the very beginning that there is no privacy. That is, my opinion about a student is a private matter perhaps between me and the person who wants to employ him and I might be willing to talk to an employer about a student, but when that is broadcast all over the Government, given to anonymous people all over, then it becomes a very unprivate affair. So we see there the exposure of the individual, whatever his weaknesses are, to the scrutiny of everybody, so there is this second legal issue about privacy that seems to me fundamental.

Now, the right of privacy is not spelled out in those words in our Constitution and I think there is a reason for that. I think that the reason is in the 18th century, when the Constitution was written, Government was very limited. Government was supposed to do only a few things, minor things we would say today, and I think that the idea of privacy being invaded in the way it can be today never occurred to the people who wrote the Bill of Rights. But everything that did occur to them in the way of invasion of privacy they wrote in. So what you find if you read the Constitution is that in every way in which they understood privacy then, they protected it in the Constitution.

Let me illustrate: They protected speech and expression and beliefs, and those, it seems to me, are illustrations of privacy. They protected religion and conscience, each individual's to be his own. They forbade the quartering of soldiers in houses. They protected the right to bear arms—and there are many other reasons for that, but one of them is the man with his rifle in his home. They protected people against the search of the person or the search of the home without a warrant and in unreasonable circumstances.

They protected people against being forced to incriminate themselves by any official body. They protected people against cruel and unusual punishment, thus in effect protecting the body against invasions that were deemed unreasonable at that time.

When they got through with all the above, they protected all the other rights of the people not enumerated in the Bill of Rights and not specifically handed over to Congress by a particular part of the Constitution.

I say, myself, that that is privacy as they understood it. That is all of the invasions of privacy that they knew of in their time, and had they known of these, it seems to me they would have dealt with them the same way.

Now, all of us know that our Constitution was framed to exist over the ages and thus has always been interpreted to deal with the problems of our times, of the times that occur and it seems to me, therefore, that there is a constitutional right to privacy which does exist and it includes the right not to be defamed by anybody in official position.

I might say there that Justice Brandeis was the author of the famous original article on the right to privacy and he was talking about this very question of defamation, of slander. He was talking, in fact, about newspapers, and whether they could invade privacy and the right which has become established in part as a result of his article back in 1901, which he found in the Constitution, as I find in the Constitution, began with the idea of defamation, the idea that we deal with today.

Well, that really takes me to the question of, is this necessary? That is, do we have to have this? Is this another example of something where we all regret it, but we have to have it anyway? Must we have the data center to save money, to save time or to save the Nation from some danger?

Necessity, I would like to point out, is a relative word. Things are necessary in one time and not necessary in another because necessity involves a balance.

I might tell my dean that I have to have three secretaries and a thick carpet and an office twice as big as I have and I would say that is necessary. "I can tell you why I need each of those secretaries." But the fact is, in our school other things are necessary too; he figures out how necessary and what is needed in other departments.

Necessity is a budget. Necessity involves a choice among things. I tell you I have to be somewhere at such and such a time and if something more important happens, I don't have to be there.

So when they say a data center is necessary, they are saying it would be useful. They are saying, "We could use one." They are saying, "If we had one, it would do the following useful things" and they list them: save money, save time, and so forth. That tells you it would be useful, but it doesn't tell you it would be necessary.

When we decide whether a thing is necessary, we have to figure out the losses. What would be taken away by this proposal?

There, it seems to me, we come to what all of us realize this proposal is going to do something to the character of the American people. The question we have to think about is what is going to happen to the character of the American people and how serious is it?

We have already had mention this morning of the great American idea of "beginning again," starting anew, getting a second chance, and that is something we would lose by this. We would have a situation in which nobody got a second chance, no matter how young, no matter how foolish, no matter how easily explained the circumstances; we would establish a doctrine of no second chance, no forgiveness.

One life, one chance only. That seems to me very different from the American dream.

We would have, in addition to that, a waste of people, a human waste. We have a lot of people in this country but I don't think we can afford to waste them and many people, including some saints, have done wrong early in their life and then lived to be worthwhile people.

Some of the best people are people that did something wrong and then went on in the light of their understanding of those wrong things, to do something right. And I can think of a lot of great people in the world who, if they had been stopped because they once took the wrong path, they would have been lost to us. They would have been some of the greatest religious leaders, some of the greatest political leaders that we have. We ought not to build into a machine anything that will keep us from having the benefit of those who once made a mistake in their lives.

A third thing that we will have, it seems to me with this center, is a nation of people who are afraid. One way I tried to prepare for being a witness here was to talk to people in New Haven. A man in the dry-cleaners listened to the best I could describe the Federal data center and the first thing he said was, "Why, that sounds just like Russia. We are going to have a number instead of a name over here pretty soon and it is not going to be very American."

I said, "You put it very well. Can I quote you?"

He said, "No, I don't want to get into trouble. Don't say a word."

Then I thought, there it is. They don't even have the center yet and this man who just works in a drycleaning establishment in New Haven is afraid. He doesn't want his name mentioned to those anonymous people in Washington who might write it down and might make some trouble for him someplace, somewhere.

If we have this center, it seems to me we will increase the number of people who are afraid and we will begin to lose what we like best in the American character, people who are willing to speak up, people who are willing to do things and not look around over their shoulder to see who is looking on.

When I spoke of the Constitution and the right of privacy and the right of confrontation, the right to be present at your own trial, I would like to emphasize that I am speaking of a minimum when I speak of the constitutional law. The Constitution represents not the maximum protection of the individual, but the minimum, and in this day and age in which we have so much pressure to invade the individual's life, I think we shouldn't stop with the Constitution; we should have affirmative laws to protect privacy more than the Constitution does.

It has always been our tradition to pass more laws than the Constitution provides because we know that is only a beginning.

While I think that hearings like this—that awareness of this problem are valuable, I believe in laws. I believe that real protection in this world comes not from people's good intentions but from the law and, therefore, I would like to see some laws on the subject of gathering of information like this.

I would like, in the first place, a law that would prohibit Government agencies from asking some kinds of questions at all. Some questions are either so personal it is nobody's business or so close to the constitutional area of religion and free speech that it is nobody's business in a constitutional sense.

Some information I would say is nobody's business at all. If you don't know that, you just have to hire them without knowing it. It is just too bad, because we don't ask people about their personal lives, in some respects, no matter how much we would like to know.

Now, secondly, I would say that there is some information that may be needed at one particular place. It may be needed for one particular man or agency, but I would say it is to go to no one else. There are other kinds of information that I would restrict only to the place where it was originally needed and I would have it sealed or destroyed when its usefulness had ended. And I would not assume that information that is useful for one purpose ought to be handed out to anybody else for other purposes.

Every time that information is required, I think it has to be justified all over again. It seems to me a very terrible society—and I mean terrible in the sense of frightening and terrifying—where information given to one man becomes information available to all.

I might very well be willing to talk to a Congressman who wanted to hire a confidential assistant, about some student, but I would never talk to the Congressman if I thought he would tell everybody in the world what he had heard from me. So I think you need a law to protect people against that.

I think you need a third law that would see to it that people have a chance to know what has been said about them and to rebut it.

I am constantly being promised by these letters and references I get that everything I say will be kept confidential, kept from the student.

Why am I entitled to that kind of protection, me as a private individual? Why should I be given the privilege of saying what I please about students and not have to account for it?

Now, in my role as a teacher, I have to account for it. If I give a student a bad mark, I have to see him the next day in the hall and I have to say "I'm sorry. You didn't do very good work." I have to face that and I have learned to face it. Sometimes it's the person you might like best in the class, you feel quite badly about it, but that is something you have to do.

Why should I be allowed to say something bad about the same student in private and get away with it? I don't see that I am entitled to any such privilege and I don't want that kind of privilege, but every one of these letters assures me that anything I say will never be known to the person involved.

I would like to make the accusers in this situation responsible for what they say and I would like to give the individual a chance to explain, to give his side of the story. That way we won't petrify this information. It seems to me that a proposal like this one for a Federal data center is an example of the process by which government "just happens."

I don't think there are bad people in government who want to destroy the privacy of the individual. I think there are very well-meaning people in government who follow their own jobs out to their logical conclusion and see in this certain advantages for efficiency and so forth, and don't see any more. So a center like this is all too likely just "to happen." It is a very great thing, I think, that this subcommittee is trying to make people stop and think. It is trying to make people, it seems to me, ask themselves, "Is this a proposal we really want, or is this, on the other hand, a rather incredible thing in this country that we should have a proposal to have a file on everybody?"

I guess all that I wanted to say today is simply that we should not permit it to happen. We should have it happen only if we want it to happen. For myself, to those who say that this particular center is necessary, that we have to have this, I would say that what we have to have is a nation of people who are independent and unafraid, and I think that is the only real necessity.

That is the end of my statement. Thank you very much.

Mr. GALLAGHER. Thank you very much, Professor, for your very enlightening statement. It certainly has touched all the bases of our concern.

In behalf of the committee, Professor Reich, we want to thank you for a wonderful contribution to this dialog. I think you have set the basis for some in depth thinking into a problem which disturbs all of us. For your splendid presentation, we are most grateful.

You have touched on many of the things that have concerned this committee in the past. The question of wrong information or bad information becoming institutionalized in Federal dossiers is presently one of the weaknesses of our personnel structure here in Washington. We have had many illustrations of this thing happening now.

The frightening prospect is if this is put into a computer and lasts forever or is made permanent, as you have mentioned.

Only the other day I had a boy graduate from school, an honor graduate, a man about 6 feet 4, who thought he ought to be a Marine officer. It would seem to me he would be the ideal type we want especially since he wants to do his part for the country, being fully qualified physically and academically to be an officer. But shortly before he was to be sworn in, a central traffic data bank turned up the fact that he had 4 parking tickets when he was 17 years of age, and this disqualified him. It showed a certain instability, even though he is now 23 years old and appears to have the right drive to defend his country and get shot at. I do not think that parking tickets ought to be a prohibition to this opportunity.

The frightening part of it was that he subsequently applied to the Navy, who recruited him as officer material, and then it turned up that he was a reject of the Marine Corps and they wanted no part of him.

So you have put your finger on the heart of the problem. We are worried about the dropouts of today, but I am worried about the computer rejects of tomorrow. Within this lies the fact that there will be no opportunity for repentance, rehabilitation, forgive and forget and go on to a happy life, and we are about to change one of the basic and fundamental structures upon which our environment is built.

You have made three suggestions about how we could protect certain taxpayers in the event we get on this path of a central data system. Would you mind outlining those again for the record?

Professor REICH. Surely. Each of these has to be spelled out after hearings, study, and so forth.

First, I think there are questions which nobody has a right to ask any citizen in this country at all—period. I think there are kinds of information that are just plain nobody's business. That is an old-fashioned phrase, but I think it is a good one to remember once in a while. There are just things that are no one's business, and as for those questions, for example, about personal life, habits and ideas, it seems to me we should have a law and eliminate.



To be a little bit more specific, I think we all agree that a question about a person's religious belief is a question that no one in this country has a right to ask anyone else—period. The Constitution says it, but I would like to see the law make it clear to every employing officer in this country as well. That is an easy illustration. Things that go on between a man and his wife and between a man and his children are nobody's business. I think we could think that through and could come up with a list of information. I hardly have to mention it, except that people ask those questions.

Mr. GALLAGHER. You are absolutely correct, because the experience of this subcommittee has demonstrated this has become a way of life, especially with the personality questions proposition that was rapidly becoming a condition precedent to Government employment. Fortunately, through Executive order, we have gotten out, at least temporarily, the questions relating to sex life and private beliefs. I agree with you, perhaps a law is needed, because while Government employees are no longer required to do this, still this exists in private industry and anywhere else that a personality-type question can be asked.

Mr. CORNISH. May I add something at this point, Mr. Chairman?

Professor Reich, if we do enumerate a list of types of questions that cannot be asked or areas which cannot be inquired into, do you not then think there is perhaps an implication that questions outside of these enumerated ones would be all right to ask?

Professor REICH. There is a little danger of that, it is true, except at the end of the law you could say that the listing of these categories here should not be taken to imply that all other questions are authorized. That is done in the Constitution as well. Indeed, the Constitution at the end of the Bill of Rights says the enumeration of certain rights here is not meant to disparage others that are reserved to the people.

Since we are told they now ask everything, it seems to me the danger you suggest is less dangerous than the danger we now confront.

I would like to add to what I just said that I, myself, would not apply the law merely to Government, but I would apply it to those corporations that Congress has power over; namely, those that are in interstate commerce. In other words, I myself think it is just as bad for a private company to do this as for the Government. Congress has power to regulate wages, conditions of employment, and many other things, in these industries, and I think they should regulate this as well if it is now being abused, and I gather it is. So, I would have a law that would be public and private in its application.

I would try to keep the categories somewhat broad in order to avoid the danger you have mentioned.

There is a second kind of law that I had in mind, and that is limiting information to the original purpose intended. It may be that one ought to disclose some aspect of his financial activities to the Internal Revenue Service, but that does not mean that that is relevant to every other person who wants to know about this individual. There are some jobs that are extraordinarily sensitive and confidential in their nature. We do not even have to talk about the area of defense. We can talk about a job in which you are a personal and confidential assistant to a high Government official. Those jobs require some kinds of information that other jobs do not.

So, I have tried to think of the categories of information, of inquiry, and limit the distribution of information once obtained. I think the principle I would have is I would start with the idea that information goes only to the person who is authorized and to no one else, unless there is a specific exception. In other words, my principle would be no distribution beyond the original recipient of information without an exception.

I take it today we have the opposite principle. Distribution is allowed unless there is a prohibition. It seems to me that is the wrong approach.

Mr. HORTON. On this very point, I would like to put into the record at this juncture the article by Mr. Macy in the Saturday Review of July 23, 1966.

(The article referred to follows:)

[From Saturday Review, July 23, 1966]

#### THE NEW COMPUTERIZED AGE—4: AUTOMATED GOVERNMENT

*How Computers Are Being Used in Washington To Streamline Personnel Administration—To the Individual's Benefit*

(By John W. Macy, Jr.<sup>1</sup>)

In any examination, whether in high school or college or in a civil service written test, it has never been considered cricket to show your paper to anyone else. In these days of automated examinations this same rule may be carried to the ultimate extreme: the only eyes that ever fall upon an applicant's civil service test may be his own. Even though the test may be sent across the continent, graded, and compared with the papers of other competitors, and even though the applicant may be hired and enter upon a lifetime career largely on the basis of this test, nobody but him need see it after he completes it.

This is one aspect of automation that bids to revolutionize personnel management in the Federal Government. Some may regard this feature as depersonalizing. But the truth is that mass examination scoring never was a highly personal activity. The automation of much personnel work of a clerical type may well serve to increase the personal attention managers can give to problems requiring human attention.

Automated examining techniques used by the U.S. Civil Service Commission may be both more advanced and more limited than the general public realizes. During fiscal 1966 the Commission's computer automatically scheduled more than 700,000 applicants into 1,000 examination points throughout the Nation, computed the scores of those who took these nationwide examinations, and notified applicants of the results. On the other hand, these high-volume figures deal only with nationwide written examinations. In many instances, persons who apply for positions are not tested, but rather are evaluated by a team of experts in a specific occupation, and are graded solely on their previous training and experience.

For one of the 700,000 persons who applies for an automated examination, the initial action on his part is simple and easy. He files only a small card form. In due course he receives an admission card, telling him to report at a specified date and hour at an examination point convenient to him. His examination has been scheduled by machine, and the time and location have been printed automatically. In the examination room, the competitor marks his answers to the questions by shading the appropriate block on a set of test-answer sheets. When the sheets are returned to the Commission, computers then take over the next steps. Their output even includes a letter to the competitor notifying him of the test results.

The notification letters roll out of the computer in one long sheet, are mechanically separated, and are finally stuffed into mailing envelopes virtually untouched by human hands. Not only is this process immensely faster, it is more accurate and requires substantially smaller expense than processing by hand.

<sup>1</sup> John W. Macy, Jr., is Chairman of the U.S. Civil Service Commission.

In addition, the computer is programed to check the validity of test results, to prepare studies showing how different groups of applicants performed on various sections of the examination, and to assist in establishing appropriate passing scores.

Automatic data processing has been applied to personnel management in the Federal Government for only about 6 years, yet the roots of the cybernetic revolution in government extend back to the 1880's. In fact, there is reason to think that the entire development of automatic data processing was initiated by an invention of a young Census Bureau employee appalled by the paperwork of the 1880 tabulation.

Herman Hollerith was a young engineer working on the 1880 census. Seeing a need for something better than handwork on the mass of census statistics, he put together a tabulating machine that he called his "statistical piano." It was somewhat reminiscent of a player piano, in that it used a roll of punched tape to feed instructions into the machine. People who, then as now, condemned the civil service for a lack of imagination and innovation, must have been looking the other way. Even the inventor may not have realized what he was starting, but in the 1890 census Hollerith's device was credited with saving 2 years of work and \$5 million. Later it became the foundation for a phenomenal business—the company now usually referred to by the initials IBM.

The Government also pioneered in the development and use of electronic data processing. One of the first completely electronic computers ever built was called ENIAC, for electronic numerical integrator and calculator. It was produced by the War Department and the University of Pennsylvania, working together in 1946 to solve problems in ballistic research. In 1951 the first commercial computer, UNIVAC I (universal automatic computer), was installed in the Census Bureau, some 3 years before a private company put a UNIVAC into operation. The Government received good value from its investment in UNIVAC I, running up more than 73,000 hours of operational use on the machine before retiring it to the Smithsonian Institution in October 1963.

When first developed, the digital computer was used merely as a large and very fast calculating machine, or for complex accounting and statistical purposes. In Government, priority was given to its employment in the primary mission of the agency by which it was used. By the early 1960's, however, the Department of Agriculture was using computer facilities for centralized personnel management data processing purposes. Its MODE (management objectives with dollars through employees) system is a large-scale centralized personnel recordkeeping and reporting operation, utilizing a computer in New Orleans. In addition to records and reports, the system computes the pay checks for Agriculture's 100,000 employees throughout the Nation.

The Veterans' Administration, with 156,000 employees, was the second large agency to install a centralized, automated personnel system. This system, called PAID (personnel and accounting integrated data system), operates at Hines, Ill. PAID encompasses general personnel management statistics and reports, career development and training records, a file on employees' length of service, payroll information to permit computation of checks by the machine, and information on the authorized number of positions as compared with the number of employees on the rolls. The system also contains a "suspense" file of personnel matters to be brought up on certain dates.

Twenty-two agencies of the Government now have automated personnel systems covering 1,500,000 Federal employees. Systems covering an additional 500,000 are being developed.

The Civil Service Commission first entered this field in administering the Government-wide retirement system. Through an automated procedure, 750,000 retirement accounts are maintained with an annual increase of 45,000 new annuitants.

Three years ago a 5-percent increase in all current annuities was authorized by Congress. This necessitated recomputation of the annuity for every person on the retirement rolls. The last time such a task was required it took months. The added workload was augmented by a stream of letters from Congressmen, justifiably wanting to know why their constituents were not receiving their higher retirement checks, but in 1963, thanks to the wondrous capability of the computer, 630,000 annuities were recomputed in just 10 days and checks started flowing out before complaints and inquiries began pouring in.

The system is now being used to compute deductions for medicare payments for those annuitants who are not receiving social security benefits. A recently

completed management study of retirement and insurance operations indicated that over the next decade more than \$3 million can be saved through further automation.

Increased automation is to be expected. But it is time to ask searching questions about these systems and what they should be doing for us. Automated personnel systems put into use during the past few years are basically record-keeping and reporting systems. But it is a serious mistake to think of personnel offices primarily in terms of records and reports.

Personnel management is principally concerned with finding the best qualified people to fill vacancies, insuring maximum utilization of manpower resources, improving working conditions and thereby improving work—and providing equal employment opportunities to all our citizens, not only at the point of entrance into the service, but through training, promotions, and full career development. Seen from this perspective, automation of personnel operations is just beginning.

As we advance, the question facing us is this: Which parts of the job can a computer do better—and which can men do better? We know that a large part of management is actually clerical decisionmaking, though we have often dignified it in the past with the word "judgment." It requires the identification of relevant facts and the selection of predetermined action on the basis of those facts. This a computer can do beautifully.

In scheduling civil service examinations, for example, our computer makes "decisions" of this kind by the thousands. Why should the time of a man or woman be devoted to such work with less accuracy and little satisfaction? We have other work for men and women, in which they can do a far better job using the huge data resources of the computers. This work involves decisions on personnel planning, the matching of men and jobs, the forecasting of manpower needs, and the important decisions of career-planning.

For proper decisions in these areas we must have integrated information systems. This will require the use of information across departmental boundaries. It is here that current efforts to standardize symbols and codes will pay dividends. Direct tape-to-tape feeding of data from one department to another may become common. These systems will mesh well with developing plans for an executive-level staffing program which will be designed to locate the best possible man for any given top-level assignment, no matter where in government he may be serving.

The computer's ability to search its perfect memory and pick out records of individuals with specific characteristics has been applied in the search for candidates for Presidential appointments. A computerized file containing the names and employment data of some 25,000 persons, all considered likely prospects for federal appointive positions, is searched electronically. This talent bank, with its automated retrieval system, broadens the field of consideration for the President in critical decisions of leadership selection.

Throughout the Government, one of the great responsibilities is to provide true equality of opportunity in employment. To know where we have failed to provide it, where we have succeeded, and how best to plan, we need a multitude of data. Through head counts we know only that a certain number of Negroes, for instance, were on the rolls in certain grades at a certain time in the past, and now we can count that there are fewer or more. But these data do not reveal whether the people in certain jobs came from lower jobs or from outside the Government. They do not assist us in recommending training or evaluating it. They fail to give us the management information required to do a conscientious job of creating conditions that will make a reality of equal opportunity.

To obtain additional and more accurate information, a new effort has been initiated in this area. By means of a voluntary racial designation prepared by employees themselves after employment, reliable information now can be fed into computers where it can be confidentially stored and used.

Most Federal managers need more knowledge of computers in order to best use their capacities. With this in mind the commission last year established an ADP Management Training Center in Washington. More than 2,300 Federal employees have attended its sessions.

There seems no doubt that increasing use of computers in Government, accomplishing many of the clerical tasks by machine, will affect the skill requirements and the "occupational mix" of government service in the future. The Civil Service Commission has made an extensive study of this question, and is giving it continuing attention. Employee displacement has not been extensive; with intelligent planning an agency can prevent hardship for the employees affected.

An outstanding example is furnished by the Internal Revenue Service, which has done an exemplary job of minimizing the impact on employees in its extensive ADP conversion program through advance planning, and intensive retraining and placement efforts. This is the kind of personnel job no computer can handle.

This seems to me to be the answer to those who fear that computers will de-emphasize humanity. Far from it! By removing the clerical decisions and the mass of paperwork details the computer may well free the mind of man for more worthy use.

Already it has heightened the need for imaginative and innovative managers who can grasp ideas, think in broad, philosophical terms, and apply such terms in decisions relating to public welfare. It has forced a finer degree of quantitative precision in executive judgment. It has liberated the manager to give his mind to greater scope of creativity. Rather than degrading the worth of the human being, the computer has placed a premium on man at his best.

Mr. HORTON. On page 25, Mr. Macy in his article says, picking it up and perhaps it may be out of context:

In scheduling Civil Service examinations, for example, our computer makes decisions of this kind by the thousand—

Meaning judgments that are clerical decisions about whether people are qualified or not.

Why should the time of a man or woman be devoted to such work with less accuracy and less satisfaction? We have other work for men and for women in which they can do a far better job using the huge data resources of the computers. This work involves decisions on personnel planning, the matching of men and jobs, the forecasting of manpower needs, and the important decisions for career planning.

On this very point that you were making with regard to stopping the information at a certain level and not letting it get across into other departments, Mr. Macy continues and says:

For proper decisions in these areas we must have integrated information systems. This will require the use of information across departmental boundaries. It is here that current efforts to standardize symbols and codes will pay dividends. Direct tape-to-tape feeding of data from one department to another may become common.

This is Mr. Macy talking with regard to the role that he has as Chairman of the Civil Service Commission.

These systems will mesh well with developing plans for an executive level staffing program which will be designed to locate the best possible man for any given top-level assignment, no matter where in Government he may be serving. The computer's ability to search its perfect memory and pick out records of individuals with specific characteristics has been applied in the search for candidates for Presidential appointments.

This is Mr. Macy saying it already has been used.

Mr. GALLAGHER. Patronage of the computer.

Mr. HORTON. And I think I would rather have the patronage system than this.

A computerized file containing the names and employment data of some 25,000 persons all considered likely prospects for Federal appointive positions, is searched electronically.

This is not what we are going to do 5 years from now, but this is what we are doing now.

This telebank, with its automated retrieval system, broadens the field of consideration for the President in critical decisions of leadership selection.

Professor REICH. That makes me think of something that is almost within my own field. It seems to me to be in direct response to that.

In the field of constitutional law, one of the things that is always

amusing to the student of the Supreme Court is that over and over again Presidents have appointed men as Justices of the Supreme Court who they knew well, had seen for years, had known in government and so forth, and then been utterly surprised at how the man did the job. We have instance after instance.

One of the famous ones is Theodore Roosevelt. He appointed Justice Holmes, who had already been a judge for, I guess, 20 years, and was dumbfounded at the kind of judge that Holmes turned out to be on the Supreme Court.

The point I am making is, what is all this information really worth, because we don't have a science of knowing how a man will do a job. If we appoint a man, we have to take a chance, as the President of the United States does when he appoints a Supreme Court Justice for life. That is a pretty big chance. The fact of the matter is that probably 95 percent of this information is utterly worthless to predict what the man will do and for most purposes I would say 100 percent. If it is a place where you can't afford to take a chance—for instance, the man who is going to pilot the plane that I have to fly on this afternoon—I would like to be sure he knows how to fly. I wouldn't like to take a chance on that. I would like somebody to certify that he is a licensed pilot. There is no second chance there.

But not all jobs in this world are quite like that and not all information is that critical. That is the 5 percent that I was thinking of.

For the most part when you appoint a secretary, an assistant, a judge, or anyone else, even if you know him very well, it is a question in the future and I think that people are trying to be more certain of the future than the future allows by using this information. I don't think a computer or anybody else can tell you whether you are hiring a good man or a bad man.

I think this is an effort for certainty where life teaches us that certainty doesn't exist.

We hire faculty members and we try to find out everything that we can about them but they don't always turn out the way we thought. I just don't think that a computer would help us, or anyone else.

The third kind of law that I had in mind is one that would tell a man what was in his own file and give him an opportunity to rebut.

Now, whether in every instance you would also tell the individual who it was who said what about him—a question that we would again argue out in the case of different kinds of agencies and different kinds of settings. But at a minimum, I think everyone is entitled to know. There are these things which have been said against you. There is this kind of information that ought to be answered. Maybe a simple letter will explain it.

The four parking tickets is a good example of that. It occurred to me right away there is a possibility that somebody else was using his car. He got the tickets, but the officer never knows who parked the car and I might find that someone who borrowed my car had gotten four tickets.

In fact, I have known it to happen that you leave your car at a service station and after they get done greasing it they put it out in the street and it gets a ticket and it is their fault it gets a ticket. We have all had experiences like that. There is a simple explanation. Of course, I don't mean a marine ought to be barred from service be-

cause he has a parking ticket but, at a minimum, he should be able to explain. He should be able to say, "I parked my car because I had to go to a very important meeting. I couldn't leave before the meeting was over and the meter ran out. So between leaving the car parked for 10 minutes more and walking out when I was doing something important, I felt I had to stay."

So I am saying that many, many times people can explain things, and only the person himself knows how to explain it. No one else can explain for him.

I think this is a vital protection. The example we had from the chairman was an example of a crime, a crime of overtime parking.

In the case of things that are less than crimes—for instance, being too nervous or something like that—there is all the more reason to explain. You might say I was nervous because my wife was about to have a child that day when this man observed me. I think that is just an elementary requirement.

Again, my principle would be, we should all be entitled to know what information is in the file and have an opportunity to explain it, except in any case where the Congress decides that it is just absolutely necessary to keep it secret. I don't know if there is any such instance, but if there is, it should be an exception and it should be a rare exception.

Those are approximately three laws, just in a general sense, that I think we need in addition to the Constitution.

Mr. GALLAGHER. Then you are suggesting, Professor, that technology perhaps has outrun the law and it is time for us to bring an extensive balance back into the overall picture?

Professor REICH. Well, I agree with that. I talked to an executive of IBM before I came down here and he said a very sensible thing to me. He said, "Don't go down there and blame the machine." He said, "What is the matter is that you don't have good enough laws to protect people. The machine will do the bidding of our society. It will turn out anything you want and there is nothing wrong with computers," and indeed, I hope I haven't come down here and blamed machines. It is a failure in laws.

I said before that I didn't want good intentions, I wanted laws. I say that because, if you take a clerk in a Government office who reviews files, the clerk has to face up to this question of responsibility I mentioned before. Here is a person with something bad about them. "What should I do about it?" he says. "I don't want to get caught by having ignored this thing."

He needs a law to help him ignore something that he should ignore.

Laws stiffen our backbone. If you want the Government employees to hire the man they ought to ignore information which should be ignored, you need a law to make them know that the Government will support them if someone challenges it later. So if the boss comes in later and says, "Why did you take this man on? He has three parking tickets on his record and now he turns out to be no good. See? We should have predicted that."

The answer should be, "I am not permitted to take that information into account. There is a law here which says so." That will keep the subordinate from getting into trouble. As I say, I believe in laws.

Mr. GALLAGHER. Of course, we are running into some trouble right now where we have laws governing certain instruments which people are violating in pursuit of enforcing the law.

The committee is on a parallel course. We are not attempting to turn back progress. Indeed, even if one could, which one can't, that would be a most undesirable attempt. But we are attempting to air the need for updating our laws, and public interest in some of the things that are happening so that these laws can be passed here in the Congress and wherever else it might be useful, in the hope that the Federal Government can set some kind of guidelines.

As Mr. Packard cited earlier, there is a feeling of frustration about all this—that people feel technology is outpacing them, that they are unable to keep up with it, and therefore there is not much that can be done except resign yourself to the fact that you must answer the questionnaire, you must take a lie detector test, you must expose your entire life and stand psychologically naked before anybody who wants to take a view of what your posture is at the moment you apply for a job.

In this way and by this means you really have stimulated our own thinking on this committee here this morning.

Tomorrow we are going to have the Government witnesses who are advocating the central data bank.

Do you view this as a necessity and, if so, would you like to comment specifically on the concept of a Federal data bank for the collection of information to be used for broad-gage planning?

Professor REICH. Well, the simple answer to your question is, I do not view it as a necessity. As a matter of fact, I think it would be largely useless. Not only is it not necessary, but I think one of the things about our present-day society is that we are flooded with information that is useless to us. We have so many thousands and millions more facts than anybody can do anything with that any proposal to have more facts, or have them more readily available is suspect from the beginning. I don't want to know all the things. I can't keep track of all the things that happen in my own narrow field. Nobody else can either, and so I think we suffer from heaps of useless information. I think that often the only thing the information does is harm. I think this is a wonderful illustration of it.

I think as to the data center itself that it would very frequently do harm, that it is difficult for me to imagine when it would actually help somebody to select a good employee. I would say in one case in a thousand it might help, but I am not even sure this is true. It would not help me to pick an assistant to know what he had done wrong in high school or anything of that sort. As for how it fits into our society—one of the things you have to think about is that the Federal Government sets an example. If the Federal Government does this, it is natural for private employers to think this is the right and necessary thing to do. The Federal Government is something that most of our people look up to and if it says we have to do this, it seems to me that every businessman is likely to say the same thing, so I think it is a mistake to set a pattern like this unless you think it is the kind of country we want to have.

As I have already said, I think very emphatically it is not the kind of country we want to have. It seems to me it would create a cate-



gory of marked individuals. I will use one of my colleague's phrases when we talked about this in a faculty meeting one day. He said we would be tying a tin can around this man. All the rest of his life, wherever he went, he would have a tin can jangling along behind him. I think this is a proposal to tie tin cans around all kinds of different people, some of whom are guilty, some of whom are not, indiscriminately. It is like marking people by cutting their ears or something.

Mr. GALLAGHER. The "Scarlet Letter."

Professor REICH. That is right. That is a more American example and it is a very good example. I think it is a very good example of a proposal where it is not necessary.

Mr. GALLAGHER. The present proposition that we are considering is not necessarily a data center to select individuals for employment although there seem to be many areas where this presently exists. The present proposition is one in which the 20 collecting agencies would centralize the information they have collected, distribute it—ostensibly for such purposes as broad planning of cities and urban renewal—and much of the data could be used in the long-range planning of the Federal Government.

This is the foot in the door as we see it to the further use of central data planning. Would you accept it on the narrow base that is presently offered?

Professor REICH. I would not, unless somebody showed me how this would help the planners to plan. I work in the field of planning and I teach a course in it. I am never sure about why they need to know X or Y or Z for a plan. For instance, if you want to decide we need new housing, do we need this kind of information I was talking about today to know that we need new housing? I don't think we do. I don't think it would help us a bit. It is again an example where most of the things that we know don't help us to make decisions and they get in the way.

That is to say, I would like the planners to go out and work on getting better housing and not spend their time reading all this useless information. I would like them to get to work and I think the more they leaf through heaps of paper, the less time they will have for planning.

Unless somebody says, "We want to know how many people live five in a room in New York City," now that would be useful. That isn't information about individuals with names. If I were trying to figure out where we needed housing, I might like to know that, but I wouldn't necessarily like to know about the emotional stability of the five people in the room. I think they would have to be pretty stable to be able to stand each other, but that is all I would need to know.

I should say, sure, some information is needed, but probably not this. I would say the burden of proof is on those who propose it to show that it is valuable.

Mr. GALLAGHER. Some of our economists and our planners feel that one of the great sources, untapped, for information, is in the Bureau of the Census. That is, that it is this kind of information they would like to have access to—even though I understand last week they came up with a planning concept which showed there were 1 million Indian maidens under the age of 19 who now live in Westfield, N.J. I am not quite sure what the spinoff is, and how accurate this particular

thing is, but nevertheless, this is the kind of information that is acquired.

The economists advocating a central data bank believe that to acquire all this information and not use it is in itself a great waste. This is one of the chief justifications for the establishment of a central data bank.

What would your comment be to that proposition?

Professor REICH. Many of the statistics are already published in general. I mean, there is a great big, thick volume so the economists can find out how many people live in a particular place and they can find out general information of that sort.

I am not sure they are making very good use of what they have now in planning. That is the first thing.

Secondly, I think we seldom recognize how much planning involves decisions which can't be based upon statistics and information. A good example is deciding where to build a highway. You probably read in the paper they are always having a fight about whether a highway should be in one location or another. Those questions don't get decided by information. They are questions of whether you are going to put it through the park or whether you are going to destroy these people's homes, or whether you are going to make it accessible to a factory. They are really political choices, or value choices, and I think a great deal of planning is far removed from the science that it is claimed to be and it belongs in the area of government, judgment, politics, statesmanship. So again I mistrust the people who say we have to know all of these facts in order to plan. My experience is to the contrary.

Again, I would say if there is something they find out, like how many three-member families, how many four-member families, how many five-member families do we have, we might know how big to make apartments. That would be useful; but more than that, I question.

Mr. GALLAGHER. You are stating then that the price we may have to pay for this information is way out of proportion to the use to which the information can be put?

Professor REICH. I am saying that and I am also saying it may be useless information that will just waste people's time, in a sense.

In New Jersey I recall a great case in which they limited—they said that you had to have a certain number of square feet in your house. It was a zoning order and a planning order. No one may build a house without a thousand square feet. That was made on a lot of research showing the average three-member family needs 1,000 square feet to live in.

I say, of what earthly use was that study? People have lived for thousands of years in less space than that. Others live in larger space. It is a completely relative question or idea. It depends on who the people are, how big the family is.

It seems to me if they want to have houses of a certain size in that particular township in New Jersey, that is a political judgment and I can make it, but I think to make it on studies and research and statistics is a fallacy.

I think the Center would produce more fallacious than good thinking.

Mr. HORTON. I don't want to completely disagree with you on this, but I want to illustrate how this type of information can be valuable

in the very field in which you are involved. A lawyer's role, of course, is to represent his client and to handle his case, whatever it might happen to be and, of course, law is based on precedent. One of the difficulties that a lawyer has is to spend the time to look up all the cases that have applied to the particular given situation. You are familiar with that type of research.

Now, it is comparable, it seems to me, in the field of law as it is in this other. The planner has to have certain facts and information at his fingertips and if he had to go back and analyze all these things and spend a lot of time digging it out, the same as a legal researcher has to dig it out, that takes time away from his specialty, which is planning. And studying to find out where these lawsuits, or these precedents or these cases are is also a matter of taking up a lawyer's time. As a matter of fact, sometimes the competency of a lawyer is based upon his ability to put his finger on a case. There has been a tendency in our courts, anyway, not to regard the absence of a lawyer having that information to the prejudice of his client. There is a possibility that you can push a button on a given subject and get, in the field of law, all the cases that pertain to that particular subject so that you just have it at your fingertips.

Now, this is certainly a good use of computers, and certainly demonstrates the need for having this type of thing. Perhaps if you think of it in the context of that type of a situation for the planner—and I don't want to argue with you or try to—we could spend all afternoon here debating whether or not it is good or bad, but I think it can illustrate how a computer can be of benefit to a lawyer, or to a doctor—because there are many doctors who don't diagnose a case properly. Xerox has its home plant in my district, and I know they have been giving very serious thought to, and are working now in, the field of providing computers and information for medical and for educational purposes.

A doctor can perhaps, by pushing a button, get backup information for diagnosis or treatment or how he should handle a particular given medical situation. So it does have a lot of benefits.

Mr. Macy, who has to make these decisions—and I don't want to defend him here, but Mr. Macy, who has to make the decisions on hiring personnel, is not in the trained position you are, where you are hiring one typist or one secretary. He is hiring and the Federal Government is losing on a daily basis probably tens of thousands of people. So they have to have some means of getting information in a hurry and not having to spend man-hours in digging up this information.

So I think this is one of the points that should be made with regard to the use of the computer.

I have these points that I have tried to emphasize along with the chairman and that is the role of this committee, to try to find out how that benefit can be harnessed and still preserve that individual right that should be protected.

Professor REICH. I agree with what you say completely. I don't mean to sound like planners should go off to a desert island and know nothing, and I don't think what we are interested in is an irrational society. I mean we try to do things sensibly so, of course, there is information that will be valuable.

I think the job of the committee is to set up categories of information that are valuable on the one hand and does not invade anybody's rights in one category and have another set of information that does invade people's rights.

Precedents in law, luckily, are not people, so if you were to have all the cases ever decided on paper or computers or anything else, we would never have a person's privacy invaded if we just used the reports. That is a good example of where it seems to me no harm could come from a lawyer being able to have the cases at his fingertips. There must be many other kinds of data exactly like that. I would be very happy to have a computer do my legal research for me. I would welcome it and I hope it comes very soon.

Mr. HORTON. They are working on it.

Professor REICH. I know they are, but they are not working fast enough.

The thing we want to do is to focus on the areas where the individual is in danger, where he is going to be scared, where he is going to be scared with reason and not just put our heads in the sand in general. I hope that was clear from what I said.

Mr. HORTON. I think your testimony before this committee has been very helpful because you have set out some basic areas in which we should be very concerned. I think you have sort of steered us through some aspects of this problem, especially on the constitutional law question. I think it has been very helpful to the committee.

Mr. GALLAGHER. If you did have an opportunity to submit a short brief to the committee outlining your proposals, we would be very thankful, especially as to the basis for regulating interstate corporations and for the regulation of problems relating to State governments and the county and city governments, the overall proposal—how you would arrive at, for instance, a proposal to limit intrusive questioning and enforce this at the local level.

If you had an opportunity to give that some thought and submit a brief to the committee, we would be very pleased.

Professor REICH. I think I will be able to do that. I will at least try. I will try to say some of these things I have said here in a shorter fashion and more precisely.

Mr. GALLAGHER. You can say them in writing just the way you have said them here. We are very appreciative.

Mr. CORNISH. In regard to your second suggestion for a law, I am concerned about the individual citizen who submits information to the Government for a specific purpose. By way of example, let's take the person who submits information on his income for one purpose—and that is to have his taxes computed for him.

Do you feel there is any element of contract in this thing? In other words, the Government asks you to provide your income data to compute your taxes and therefore they should not use it for any other purpose, or is there in a sense some informal or perhaps even legal contract that exists in a situation like that?

Professor REICH. I would call it a moral contract. I would be glad to use that notion because I think that it is a sort of an understanding by the citizen of what is fairly expected of him.

Mr. CORNISH. You did not use the word "contract" in your discussion of that. I wanted to know whether you did feel there was an element of contract in such an arrangement.

Professor REICH. I would not use it in its legal sense because the Federal Government, which is sovereign—I am just being a technical lawyer for a minute—does not make contracts with its citizens. It may ask them for what it needs and whether they like it or not they give it. That is the nature of the sovereign, to make laws and people have to obey. I would prefer to call it a moral contract of what citizens understand they are being asked for. This information is being asked for a purpose and they agree with the purpose and they agree it is necessary or they would not agree to its use for other things. In that sense I think it is a very good phrase because I think it implies what most of us feel.

It implies the same thing and again this is not a technical use of the word "contract" but if you call me up in 3 days and say, "So-and-so has come in here and is looking for a job. Is he a good man?"

You don't say so but there is implied in the call, a notion that I am helping you to find a man. I don't expect you to repeat what I say to everybody down the street. That is not a contract either because we did not exchange money and so forth, but it is an understanding. I would be kind of upset if I then saw that you told the press or somebody else what you have heard from me.

I think while we do not want to use the technical word "contract," I think that people sense their Government is dealing fairly with them is a very important thing. People ought to believe their Government is treating them fairly and the tax case you gave is a very good illustration of that. I will stick with the word. Glad to use it.

Mr. CORNISH. That is all.

Mr. GALLAGHER. Mr. Romney?

Mr. ROMNEY. No questions.

Mr. GALLAGHER. Mr. Forsyth.

Mr. FORSYTH. Can I make just one observation? I think you said the economists are not making good use of the material the Census has now. This very argument has been advanced by the economists as a full reason behind this data bank. Somewhere we have to get our definitions a little sharper to find how these go together or conflict.

Professor REICH. I would like to know what information they did not have that they now need and why they need it.

Mr. FORSYTH. Do you have Mr. Dunn's report, by the way?

Professor REICH. No, I do not.

Mr. FORSYTH. You probably ought to get that before you write your brief. It might be valuable.

Professor REICH. I think it is a question you can answer only in the concrete, and planning covers everything in the whole country.

Mr. FORSYTH. It is too broad to discuss.

Professor REICH. Should we have a train that can go from Washington to Boston in 2 hours? That is a planning question. I do not know what we need to know to answer it. I would say "yes," offhand, but maybe the answer is "no."

I think we would agree in most instances that it is important to get down to specifics about that.

Mr. FORSYTH. That is all.

Mr. GALLAGHER. Professor, while it is not necessarily germane to this hearing, there seems to be a climate of concern now existing in the

country as to various matters relating to privacy. We appear to be on a drift course to some extent on questions of wiretapping, bugging, invasions of privacy through all sorts of means, credit evaluations. We find overzealous law enforcers breaking the law to enforce the law they are sworn to uphold.

I think that there is quite a fear now existing that the Government may not be treating its citizens fairly or that the citizen is overpowered and must resign himself to the fact that the rules are now different.

While the Magna Carta was not written just for the king's men it appears that the king's men may well be rewriting the Magna Carta, to make it easier to collect taxes, or whatever the problem may be.

Would you care to comment on that?

Professor REICH. I agree with the thought. It seems to me that we should realize these things usually happen because people are trying to do their job, trying to do it too well. I don't hold to the theory that bad people do these things. I think good people do. The policeman tries to do his job and a civil service man tries to do his job. All of us require laws to remind us of all the other things in society that matter besides our own jobs. That is what a law is. It is a statement of something more general than your own personal concern.

In other words, I might see a diamond ring on the table and I would like to have it, but a law reminds me that other people have interests in the diamond ring. Someone else owns it and someone else cares about it.

A law is to give you a sense of something beyond your own task. I think that in a society like ours privacy disappears just because there are so many people. We are all so crowded together that what was taken for granted in the old days now is threatened by the simple fact of modern life. What we did not need laws for before, we do need laws for now.

When a man could have a quarter section of land and a house on it, he probably did not need a law to protect his privacy. He was probably lonely most of the time and would like to have his privacy invaded. As the times change, you begin to need laws where you did not before. I think that the country is going to keep on growing. It is going to get more crowded and what was once taken for granted is now a precious thing to try to save, try to preserve. I do not think that life would be worth living without some space left for the individual and so I think it is the job of a Congress that wants to plan for the future to look ahead to preserve the values that are threatened by changing times.

I see this as a job to keep our laws up to date with the conditions of modern life. I think that is what this subcommittee is doing.

Mr. GALLAGHER. On behalf of the subcommittee, Professor, we want to thank you very much for the splendid contribution and for stimulating our own thinking, and also for the valuable thoughts that you have contributed to this problem and to reassuring us that by updating our laws, in your opinion, that the world will still be a very desirable place to occupy and this country can, once it brings a sense of balance to fast-moving technologies, be a very desirable place to continue life.

I think that one of the main problems—and I certainly agree with you—is that we must update our laws in view of the rapidly changing complexion of our environment. For your contribution to our enlight-

enment to what we hope will form the foundation for the updating of our laws, on behalf of my colleagues I want to thank you very much.

Professor REICH. I thank you very much for the chance to appear here.

Mr. GALLAGHER. The committee will stand adjourned until tomorrow morning at 10 a.m. when we will hear the Government witnesses, Edgar Dunn and others, who are the originators of this concept. At that point we will question them on the proposal of the Central Data Bank.

The committee stands adjourned until 10 a.m.

(Whereupon, at 1:37 p.m., the committee adjourned, to reconvene at 10 a.m., Wednesday, July 27, 1966.)

## THE COMPUTER AND INVASION OF PRIVACY

WEDNESDAY, JULY 27, 1966

HOUSE OF REPRESENTATIVES,  
SPECIAL SUBCOMMITTEE ON INVASION OF PRIVACY  
OF THE COMMITTEE ON GOVERNMENT OPERATIONS,  
*Washington, D.C.*

The subcommittee met, pursuant to recess, at 10 a.m., in room 2247 Rayburn Office Building, Hon. Cornelius E. Gallagher (chairman of the subcommittee) presiding.

Present: Representatives Cornelius E. Gallagher, Benjamin S. Rosenthal, and Frank Horton.

Also present: Norman G. Cornish, chief of special inquiry; Miles Q. Romney, associate general counsel, Committee on Government Operations; and John Forsyth, special minority consultant.

Mr. GALLAGHER. The subcommittee will come to order.

The first witness we will have this morning will be Raymond T. Bowman, Assistant Director for Statistical Standards of the Bureau of the Budget, and the officer with the direct responsibility to make recommendations for the establishment of a Central Data Bank. Mr. Bowman?

Mr. BOWMAN. Might I ask Mr. Paul Krueger, who is also from my office, to come forward with me.

Mr. GALLAGHER. Yes. Will you please proceed, Mr. Bowman?

**STATEMENT OF RAYMOND T. BOWMAN, ASSISTANT DIRECTOR FOR STATISTICAL STANDARDS, BUREAU OF THE BUDGET; ACCOMPANIED BY PAUL KRUEGER, ASSISTANT CHIEF, OFFICE OF STATISTICAL STANDARDS, BUREAU OF THE BUDGET**

Mr. BOWMAN. Thank you, Mr. Chairman.

Mr. Chairman and members of the committee, I welcome the opportunity to appear before this committee to discuss problems of possible invasion of privacy which might be involved in the establishment of a Federal Statistical Data Center. For some time the Bureau has been investigating the general problem of statistical data storage and accessibility.

Electronic data processing has revolutionized the methods of recording statistical information so that such data can be tabulated and used as may be required for statistical analysis. We want to be sure that we are making effective use of new technological developments because: (1) we want to bring all available statistical information to bear on problems which confront the Nation, and (2) we want to hold down the burden of statistical questionnaires on respondents. While making



the greatest possible use of our statistical resources we must preserve the confidentiality of individual responses to statistical inquiries which we have always recognized as the foundation of our Federal data collecting activities.

A committee of the Social Science Research Council discussed the problems and prospects for improvements in statistical data storage and access with me and with members of my staff on several occasions in 1963, 1964, and 1965. It prepared a report to the Social Science Research Council which highlighted certain of the problems of accessibility of statistical data and made proposals for setting up a National Data Center as a way of dealing with these problems. The Bureau of the Budget's Office of Statistical Standards assisted the committee of the Social Science Research Council in assembling certain materials for their report. A copy of their report has been provided to this committee and has had some circulation among interested persons in Government and the academic community.

In order to examine some of the problems of a Statistical Data Center more closely, the Bureau of the Budget employed Dr. Edgar S. Dunn, Jr., as a consultant, to study the feasibility and advantages of a statistical data center and to prepare a report setting forth his conclusions. Dr. Dunn became conversant with many of the problems of data storage and access in the Federal Government when he served as Deputy Assistant Secretary of Commerce for Economic Affairs. He has also had a longtime interest in these problems as an economist and statistician. He has studied many of the aspects of data storage and accessibility as a member of a committee of the American Statistical Association set up for this purpose. Recently, he has been chairman of that committee. When he was a Budget Bureau consultant arrangements were also made for him to consult with technical staff of the Bureau of Standards. His report has been furnished to this committee and has been available to interested persons both in and outside of Government.

At the present time a task force appointed by the Director of the Bureau of the Budget is considering "measures which should be taken to improve the storage of and access to U.S. Government statistics." The task force members are Carl Kaysen, chairman, Harvard University, Cambridge, Mass.; Charles Holt, University of Wisconsin, Madison, Wis.; Richard Holton, University of California, Berkeley, Calif.; George Kozmetsky, Teledyne Corp., Los Angeles, Calif.; Russell Morrison, Standard Statistics and Standard & Poors, New York, N.Y.; and Richard Ruggles, Yale University, New Haven, Conn. The task force has not as yet submitted a report.

I have stated that the Bureau of the Budget wishes to utilize the technological advances in data processing and handling associated with the computer to make more effective use of available statistics in meeting demands for information required to deal with today's problems. More and more we are coming to realize that the problems with which we must deal are combinations of many factors and can only be diagnosed and solved by information which relates the various factors involved. Such interrelations of information by the development of a statistical data center need not pose a threat to individual privacy if such a center is governed by restrictions which prevent the release, either within Government or to persons outside Government, of information about individuals or business units.

It does not appear that interrelated use of information can be provided for by more detailed publication of data by the individual collecting agency or in other ways when the information identified with individual reporting units is not disclosed. The only other possibility is to collect the associated information de novo for each inquiry with increasing costs and burden on respondents.

The Director of the Bureau of the Budget is specifically responsible for holding this burden to a minimum. Both the President and the Congress repeatedly remind him that they expect him to perform well in this area. Thus, we continually try to find ways of using existing records for statistical purposes. We have made some progress in this effort but believe more can be done. For example, by using the Federal income tax records as a source of statistical information, it has been possible to relieve about 1 million small businessmen of the burden of filling out census forms every 5 years. And this was done without compromising the confidentiality of a single return. So successful has this been that the Census Bureau plans to substitute tax return information for census questionnaires for yet another 1 million small businesses.

While we want to do all that we can to bring all available statistical information to bear on any problem under study and while we are continually concerned with the need to reduce duplication by making the fullest possible use of existing statistical materials, we are also vitally concerned with preserving the confidentiality of information reported to the Government.

Our Federal statistical gathering activities are expedited by the prompt cooperation of respondents. While most statistical information is based upon nonmandatory replies to Government inquiries, even mandatory replies are more prompt and accurate because of the Government's proven practice of not revealing information supplied by individual persons or business units. This is a most precious asset. We could not effectively operate our statistical system without it. A statistical data center, therefore, if organized, must maintain confidentiality as to individual suppliers of data while at the same time improving access to statistical information for statistical purposes.

Under law and regulation, information reported to the Federal Government for statistical purposes is not released in any form whereby data furnished by any particular individual or business establishment can be identified. There is general recognition that this practice of confidentiality is sound public policy. Thus, when the Supreme Court in *St. Regis Paper Co. v. U.S.* so construed the confidentiality provisions of the census law as to make it possible to subpoena a copy of a census return held in the files of a respondent, the administration supported and Congress quickly passed remedial legislation to give to the copy the same confidentiality and immunity from legal process possessed by the original.

Maintenance of this principle would be a major tenet of any statistical data center and is clearly required under present law. The Federal Reports Act of 1942, section 4(a), provides that in the event information obtained in confidence by a Federal agency is released to another agency, all the provisions of law, including penalties, relating to protection of the information from unlawful disclosure by the collecting agency are also applicable to the agency to which the infor-

mation is released. Title 18, United States Code, section 1905 provides for penalties (fine or imprisonment, or both, and removal from office) for Federal employees who disclose confidential statistical data. These legal requirements can be reinforced by providing that all information in the statistical data center—under whatever conditions collected—will not be released by the center in such a way as to disclose information furnished by individual persons or businesses.

A statistical data center, as we conceive of it, is a way to improve storage of and access to information for statistical uses. It would not have an interest in building up dossiers on individuals because statistical interests do not center on individual cases. Use of data in the files of the center for other than statistical purposes would be prohibited.

I think, Mr. Chairman, even a reading of the report we have made available to the committee would indicate that some comments about them have been out of context.

In conclusion, let me note that nothing I have said should be interpreted to mean that the privacy of individuals does not require continuing and careful attention not only with respect to information now in the possession of Government, but also with respect to the collection of new or additional information.

I have made my prepared remarks brief. I hope they covered the main points. If there are questions, I shall try to answer them to the best of my ability.

Mr. GALLAGHER. Thank you very much, Mr. Bowman.

I agree that perhaps some of the remarks have been a bit out of context, but I would hope that you would agree that it would not be beyond the concern of all of those of us who have some share of responsibility to ascertain whether or not there could be a breach of confidentiality of the names of millions of Americans whose dossiers would be in a data center. The issue is not whether or not you could legally release them. The issue, to my mind, is whether or not this data will be stored with names so people would have access to this information in a central data bank.

I am sure no one would breach a law to release such information, but the fact that it is in a centralized institution where human beings will have access to it is the chief concern.

If you would assure this body today that the identification of no individual would be in your statistical bank, I think we would be very happy to say "Go about your business, and you have met your requirements or you have met at least the elements of our concern."

Mr. BOWMAN. May I comment on your statement, Mr. Chairman?

Mr. GALLAGHER. Yes.

Mr. BOWMAN. I would not want to say that within the data center, within the statistical data center that I am talking about, there would be no identification of information with an individual, just as the Census Bureau can now identify information about a particular business firm and about a particular individual. You would not be able to use this information meaningfully unless this kind of identification were maintained, particularly by the agency which collects the information or the agency which wants to assemble it for analytical purposes.

But I can definitely assure you, Mr. Chairman, that in a data center the availability of information to the staff of the center would probably be much less than the availability of the information to the staff

of the Census Bureau or the Bureau of Labor Statistics or the Social Security Administration at the present time, because the data center would have this information in machine-readable form and not on individual schedules which any one of the members of the staff could examine.

Secondly, I would emphasize that so far as the statistical data center is concerned, there is no intention to organize the data in the center with regard to individuals; that data in different files, data on different tapes, might certainly identify individuals enough so this information can be associated together for statistical purposes, but there would be no intention and no need for the data center to organize specific records about specific firms or specific individuals so you accumulated a lot of information about individuals. There would be the need for collating this information for special, particular statistical analysis, and there would definitely be a restriction on the giving to anybody, in Government or outside Government, just as the Census now does. The Department of Justice cannot get information from the Bureau of the Census. If they want a return filed by a particular business firm, they cannot get it. It is not a matter of theory.

Mr. GALLAGHER. We are trying to protect your interest and the reputation of Census and the Department by not allowing information to get mixed up with data collected by other agencies where the law is supposed to protect the individual confidentiality and confidentiality has never been a major factor.

Mr. BOWMAN. The other agencies do not have the same law as the Census Bureau, but they all undertake the same practices and I would say, Why can't the data center be governed by law in the same way as the Census Bureau is?

I would like to make one distinction, Mr. Chairman. I do not want to make my comments now with respect to all kinds of data centers, nor that this is a grand data center for all kinds of purposes. In other words, there is no use in the Federal Bureau of Investigation maintaining a fingerprint file if they are not to be allowed to use it to identify individuals, but there is no need in a statistical data center when it is being used for statistical purposes, to reveal any information about an individual. I think we have protected it in the past among statistical agencies, not only in the Census Bureau, and we can protect it in a statistical data center. Notice that in my remarks I was very careful to center attention on a statistical data center. The reports we have had have not done this as well as they should have, for they have been more technical. I am talking only about a statistical data center.

Mr. GALLAGHER. You have narrowed it down somewhat from Mr. Dunn's report, and you are zeroing in on the statistical data center aspect of this.

Mr. BOWMAN. The Dunn report and the Ruggles report, while they are not as specific as I now am, were addressed to the same idea. They were just not careful enough in their wording. What they were thinking about and at least what we were interested in, in reviewing their proposals, was not a data center for all purposes, but a Federal Statistical Data Center. We recognize that there are needs for other kinds of data centers for other kinds of purposes, but so far as the data center I am talking about and so far as the data center that the Bureau of the

Budget has been interested in in connection with my work, it has been a Federal Statistical Data Center, not a center for other purposes.

Mr. GALLAGHER. Are you sufficiently confident to say categorically that this will never be used for any other purpose when you have the capacity of a computer to do all sorts of things in this day and age?

Mr. BOWMAN. We have the capacity to do all these sorts of things right now in the Census Bureau. We can do 75 or 60 percent of it.

Mr. GALLAGHER. Let's not talk about the Census Bureau. Let's talk about the IRS. Do you think the same pattern of confidentiality exists there?

Mr. BOWMAN. No, but let me—

Mr. GALLAGHER. You are going to mix this data.

Mr. BOWMAN. Let me make this clear. The IRS data release is governed by law. These laws can be changed. I would conceive of the Federal Data Center as having information from the IRS files in it, but the Data Center would not release that information about individuals. If it were released at all, it would be released under the conditions that now prevail with regard to the IRS and only with respect to its data and by IRS. The Data Center would not itself release that information, nor would it associate that information with anything else and release it in associated form. The kind of data center that I am talking about is a Federal Statistical Data Center which would do a great deal to relieve American business of duplicate reporting, would not reveal information about any individual or any individual business, but would make it possible to bring this information together for statistical purposes when released in statistical form. The identity of the individual would not be disclosed.

Mr. GALLAGHER. Then why can you not give us assurance the identity of the individual or the individual corporation will be eliminated before those statistics will be put into the Data Center?

Mr. BOWMAN. Mr. Chairman, for this reason: Suppose we have certain information in the data that are in the Center from the Census Bureau about individual persons. Suppose we have certain information in the Internal Revenue returns. We do not want to ask the business firms to give us information they have already given us, but we must be able to take the information that we have given to us which are not on the Internal Revenue returns and put it together with the information that is on the individual returns and save them the job of giving us additional information, and make statistical analyses which will indicate various characteristics of the economic scene.

But if anybody comes to us and says, "Give us the information about X company that were on census returns plus the information from IRS," the answer is clear. It can be made a matter of law. Penalties can be placed on the people who supervise the Data Center and operate it. Information about an individual cannot be released from the Center. It is just as clear as that. On this I can give all the assurance in the world.

Mr. GALLAGHER. Are you recommending it?

Mr. BOWMAN. Remember, we are only considering a data center now. If a data center is organized, I would definitely recommend that very clear and specific regulation—legislation if this seems to be the desirable method—be inaugurated with regard to a data center, yes. I am just as much interested as this committee is in protecting the

confidentiality of replies to statistical inquiries, just as much interested.

My life work is in the field of statistics. I am absolutely convinced that the success of the American statistical program is the confidentiality it provides respondents. I am convinced this is true whether the reply is mandatory or whether it is voluntary, because it gets us rid of going through all the weeks of sending it to the legal department in the business firm to see whether they are going to answer it or not answer it.

We have an excellent relation with the business community, and we want to protect that.

Mr. GALLAGHER. Where it is on a voluntary basis and where they know it is adequately protected and not going to be exchanged with other information, the American business community has been very willing to give you information on a voluntary basis. Will not this well of voluntary flow now dry up if they know you are going to put it in a central data bank where the IRS and Census keys might get mixed up some afternoon?

Mr. BOWMAN. Mr. Chairman, all I can say is that we have to assure them in the same way we have in the past.

Mr. GALLAGHER. They have not been very greatly assured in some aspects of some of the agencies that may feed information into the central data bank.

Mr. BOWMAN. I put in my testimony the fact that at the present time in order to save business firms the need for unnecessary reporting, the Census Bureau has been given access under regulations of law to IRS returns. Here you have made in a sense the data center within the Census Bureau. You have added to their own information this other body of information. In this case we are going to substitute one for the other, but in some other instances we will merely be using it to supplement some census information. Everyone is assured that when it gets to the Census Bureau it will not have any less confidentiality than it had in IRS. In fact, maybe they are assured it will have more.

I would hope if we organize a data center, people will be convinced that any data that are put in the Data Center that have less confidentiality in the place where it came from, may still have that less confidentiality there but it will have the greater confidentiality of the Data Center. The success of a data center for the purpose that I am interested in, for statistical purposes, is in being able to develop a data center in which that confidentiality will really be recognized.

Mr. GALLAGHER. Of course, if confidentiality does exist, you are performing a very useful purpose; but I cannot help but remember a letter I got not too long ago wondering whether or not one of the drop-outs from the IRS lock-picking school might now be in charge of the Data Center.

Mr. BOWMAN. Here we are beginning to introduce a lot of things. Here we are saying—

Mr. GALLAGHER. We are talking about things that are protected by law, and you are trying to demonstrate to me that, despite the great temptation that will now exist for people who have access to information gather by agencies other than their own, no one is going to do it.

Mr. BOWMAN. No, I am not. I am indicating how difficult this situ-

ation gets. In other words, someone has information that a certain individual committed an indiscretion in the past. We have all agreed, the testimony here has indicated, that one of the difficulties with an electronic data machine is that it never forgets it and that man may never be able to get a responsible job any place again if that information is known about him.

Now we are taking an example and saying, "but suppose the Census Bureau or suppose the data center has a person like that." In other words, this person could be employed by the Census Bureau now and he might have access to this information. He could be employed by the data center. I cannot deny that. All I can say is that penalties of law for unlawful disclosure of information have been provided, and I can say that so far as the Census Bureau is concerned, I know of no significant or other than significant instances where information has been disclosed.

Mr. GALLAGHER. Let the Chair take judicial notice that the Census Bureau has been inviolate, and now let us talk about some other things. This has not always been the case with other agencies. Now you are making a mix of other agencies that might have access to that information which we are now discussing, if it is in one central location.

Mr. BOWMAN. Have you in mind any statistical agency, information gathered under restrictions of statistical confidentiality, in which there have been breaches?

Mr. GALLAGHER. No. We are talking now beyond statistical information, even though you concentrate in this area. You fail to give me the assurance that the individual's name will be deleted from your statistics, and it would appear to me we have now placed him in a position of jeopardy unless there are adequate safeguards set up in your data bank.

Mr. BOWMAN. I agree. There would be adequate safeguards.

Mr. GALLAGHER. Is this the telecommunications type of computer that we are talking about, where you would have someone at Census communicate by wire, or whatever the mechanism might be, from the Census Bureau to the central bank for information?

Mr. BOWMAN. Mr. Chairman, I think this is an area, also, which needs very careful discussion. Our consideration of the statistical data center to date has really just begun to examine the various conditions. I cannot say anything more now than what are my own personal views. The data center that we are talking about, the statistical data center, would basically be for data that already have been collected.

Mr. GALLAGHER. The 9,000 reels. Is that what you are talking about?

Mr. BOWMAN. This is information, I presume, which indicated some of the things that might be put in the data center. We have not even discussed in detail the things that would be put into the data center. There are some things that would not be.

The point I would like to make is the data center would not take the place of each of the collecting agencies that now exist, maintaining all of the information with regard to current information. Basically, the data center would be a place where one Federal agency could go and say, "We are dealing with a problem which requires us to have information on other bodies of information than those we now have."

Historic, not day-to-day information. This could be put together for them in a single center.

This is now done, but it is done by one agency or it is going to two or three other agencies to get the job done, or it could be that outside the Government there could be a request for a particular kind of analysis which requires data from more than one agency, but basically of a historic character.

I want at this stage to say we have no idea in mind at this moment of all information collected by all agencies moving to the data center by any kind of wire system.

Mr. GALLAGHER. Is this not one of the basic fundamental problems of it all as far as confidentiality and safeguards are concerned? Should you not have a concept now prepared as to whether or not this can be done before you move into doing it? One of the problems with the system that IRS has installed where they want quick results is a quick and dirty system where it has been quickly done, where you do have results immediately, but there are no adequate safeguards installed.

Mr. BOWMAN. I know that. I believe that we have not finished our investigation yet. We believe that major economies, major reductions in work done by individual respondents can be realized without doing all of those things at this particular time so far as the Federal Statistical Data Center are concerned.

We believe that this can be done without sacrificing any elements of confidentiality which now prevail with regard to statistical data, which is illustrated by—not comprehensive but illustrated by—the Census Bureau activities. That is what we believe.

Mr. GALLAGHER. How can you make that assertion when you do not know what kind of a system you are about to install?

Mr. BOWMAN. I know what kind of a system. I do not know what its ultimates will be.

Mr. GALLAGHER. I ask you if you are going to have a telecommunications type system from a central data bank to the respondent agencies?

Mr. BOWMAN. The answer is at the present time "No."

Mr. GALLAGHER. How or what use will you make now of your computers in the various agencies relating to the central data bank?

Mr. BOWMAN. Suppose, for example, the Census Bureau now has information on the population census for 1960. It has it on machine readable type.

Mr. GALLAGHER. Right.

Mr. BOWMAN. The census records themselves now have to be kept, sometimes they are kept in Archives. The Archives is in a sense a data center but a data center in which much of the information cannot be used. If we had the kind of a statistical data center that I am talking about, the tapes for the population census would be in the center in machine readable form.

Mr. GALLAGHER. How would they get there?

Mr. BOWMAN. Transferred from the Bureau of the Census to the center.

Mr. GALLAGHER. How? Hand carried, teletype?

Mr. BOWMAN. Runner. Just a roll of tape. How many rolls there would be—



Mr. GALLAGHER. When you roll off on your own machine is it going to simultaneously roll it off in the data bank, or is somebody going to roll it over in the data center and carry it over?

Mr. BOWMAN. It could be done that way. What we are talking about—

Mr. GALLAGHER. This is a fundamental problem involved here.

Mr. BOWMAN. For example, the tapes which the Census Bureau now has from IRS are tapes that were made in IRS that are turned over to the Census Bureau.

Mr. GALLAGHER. How?

Mr. BOWMAN. Physically.

Mr. GALLAGHER. Hand carried?

Mr. BOWMAN. Yes, sir.

Mr. GALLAGHER. That is a very inefficient system. Why would you not have them remain there and use the computers that you now have without centrally locating them, if you have to hand-carry them?

Mr. BOWMAN. If we were set up to do this it might be a better way of doing this. All the technological advances will not be achieved in a moment. All I am trying to make clear is that access to the data by the center, different bodies of data, does not require any relaxation of confidentiality. That is my opinion.

Mr. GALLAGHER. You are assuring me of that but that is not what you are telling me in response to questions I am asking you, Mr. Bowman. We cannot tell, when you get down to it, whether or not we are going to have a data index on every American citizen. We cannot speak in generalities. I think we ought to get to specifics. You have a proposal here that you are about to embark on.

Mr. BOWMAN. No.

Mr. GALLAGHER. You have not really thought the problem out.

Mr. BOWMAN. Mr. Chairman, I do not have a proposal here. We are considering an issue and we may have a proposal which we will want to bring before the Congress. I am not presenting a proposal now. I am explaining to you why we have been investigating the advantages and disadvantages of setting up a Federal Statistical Data Center. I am telling you what I consider to be the major advantages.

Mr. GALLAGHER. I would like to hear the disadvantages. This is what you have not responded to in the question.

Mr. BOWMAN. No. The disadvantage, the one you have mentioned—

Mr. GALLAGHER. I am asking you how will you convey the information that you gather on a confidential basis? How will that be conveyed to the central data bank if and when you ever get around to setting up a central data bank?

Mr. BOWMAN. At the moment much of the information might be deposited in the Data Center in the form of reels of magnetic tape.

Mr. GALLAGHER. Is that compatible tape?

Mr. BOWMAN. If it were not compatible it would be one of the jobs of the Data Center to make it compatible.

Mr. GALLAGHER. What would be the cost of that?

Mr. BOWMAN. Until we estimate it in more detail we cannot say.

Mr. GALLAGHER. You gave us a \$2 million estimate. That seems to be an impossible figure.

Mr. BOWMAN. Impossible with respect to what?

Mr. GALLAGHER. In making the tapes now that you have compatible.

Mr. BOWMAN. I did not give you a \$2 million figure.

Mr. GALLAGHER. It is in the proposal. There is a \$2 million figure.

Mr. BOWMAN. Which proposal?

Mr. GALLAGHER. Mr. Dunn's, I believe it is.

Mr. BOWMAN. We have an estimate by the consultant as to what it would be possible to start a data center for along these lines. We have not accepted any of these figures nor have we actually considered all of the problems that would have to be considered. We have not even considered in detail what would be the best bodies of information to have in the Data Center in its early stages. None of these things have been determined as yet.

Mr. GALLAGHER. Why do you not forget about the whole thing until you have thoroughly thought it out and in several years come back and see where we stand?

Mr. BOWMAN. At the present time, I did not think the Data Center as a data center was in review. I thought what was in review before this committee was the ideas associated with the Data Center and the confidentiality of information or the invasion of privacy.

Mr. GALLAGHER. Are not all of these involved in a central data system?

Mr. BOWMAN. I would like to very honestly answer the question which you think I have not honestly answered, or that I have not specifically answered. That is, it is not possible as I see it to have a data center that is meaningful, even a Federal Statistical Data Center, in which the identity of the individuals for which there is information is erased.

It is possible that the fact that the Data Center has access to this information does not mean that it will be revealed internally within the Government with regard to a single individual. That is the point I want to emphasize in my testimony.

Mr. GALLAGHER. That is a very good point. That is one of the things that I have been trying to direct my questions to. What kind of a system are you going to set up in order to do this?

Mr. BOWMAN. I would hope that we will have a system designed, and we will have a proposal of this character that can be discussed as a definite proposal. We do not have it now.

Mr. GALLAGHER. You think such a system is designable?

Mr. BOWMAN. Yes.

Mr. GALLAGHER. With adequate safeguards?

Mr. BOWMAN. Yes, sir.

Mr. GALLAGHER. Secure with hardware security and software security?

Mr. BOWMAN. Yes, sir.

Mr. GALLAGHER. I hope you tell the Defense Department. They do not think so.

Mr. BOWMAN. I think it is so far as Federal statistical data is concerned.

Mr. GALLAGHER. I am talking now about adequate safeguards to protect confidentiality on transmission to protect against bugging, to protect against interception, to protect against the software problems of who is going to program it, who will have the key, who will mind the tapes, all the human factors involved.

I think the Defense Department has a very vital interest in this sort of a system. If you have a system, I would sleep a lot better at night if your people would go over and tell them how to do it.

Mr. BOWMAN. The Defense Department is talking about a problem somewhat different from the one I am talking about.

Mr. GALLAGHER. No problem is more important than the protection of an individual citizen. This is exactly what I am talking about.

Mr. BOWMAN. That is what I am talking about. The system that I am talking about for maintaining that protection, for improving it, I hope does not involve all of the things that are involved in the Defense Department's operation with regard to the communication of information.

I am very sympathetic to what this committee is trying to do. But I am trying to make my point that a Federal Statistical Data Center can be developed, in my opinion, that will protect the confidentiality of individuals.

Mr. GALLAGHER. Mr. Bowman, do you know of any system now that is set up that is fully secure with regard to radiation transmission?

Mr. BOWMAN. Radiation? I do not know a thing about radiation transmission; no.

Mr. GALLAGHER. I am now sitting in this office typing up a report that is confidential and this report, if we are going to use the computer, ends up being transmitted over to my office. I would assume if you would translate this to someone down in Census typing up statistics about a copy of some sort, this would now be transmitted down to your data bank. That would be a radiation-type transmission.

Mr. BOWMAN. This is assuming we are going to use that method.

Mr. GALLAGHER. I want to know what method you are going to use. Are you going to hand-carry? If you are going to hand-carry, it is not a very efficient use of the computer.

Mr. BOWMAN. Here in a situation—

Mr. GALLAGHER. That is what I am talking about.

Mr. BOWMAN. If you are using the most advanced method you may be subject to a threat of release of information that you do not want to release. Then I would say that in the light of the public interest you may have to deny yourself the using of the most efficient method in order that you may protect the individual against the invasion of his privacy and use a less technologically advanced method.

Mr. GALLAGHER. Like what?

Mr. BOWMAN. You just said hand-deliver the tapes.

Mr. GALLAGHER. Now, if that is so, why can you not do that now?

Mr. BOWMAN. We are doing it now.

Mr. GALLAGHER. Is it working well?

Mr. BOWMAN. It is working reasonably well.

Mr. GALLAGHER. How many users do you have?

Mr. BOWMAN. I would have to get information on that now. We are making tapes available in different ways among different agencies. We think that the idea that we are setting forth here will improve that considerably.

Mr. GALLAGHER. Let me ask you this: How many users have you had? How many customers do you have? How many potential customers do you have? How many requests have been granted? How many have been denied? This would create a demand for the central data bank.

Mr. BOWMAN. Some of the individual agencies could give me information and I will be glad to get it for you as to the demands that they have had for certain of their information. In other words, the Census Bureau has developed a sample of its population census in which in this case the identity of the individuals is lost so far as the user of the sample is concerned.

Mr. GALLAGHER. I am not asking about confidentiality. How many customers do you have for this? How many users?

Mr. BOWMAN. I could find out how many tapes were sold. At the present time it is like asking a new business how many customers it is going to have. One of the things we will want to consider is whether it is worthwhile setting up such a center in terms of the service that can be performed.

Mr. GALLAGHER. Should you not have that information already if you are talking about the need for this as a matter of efficiency?

Mr. BOWMAN. Mr. Chairman, at the present time various statistical agencies are now beginning to develop tapes with information from other agencies. We are in the process now of each agency trying to constitute itself a data center. We are trying to find out if there is not a better way of serving the needs of all Federal statistical agencies just as one group by having the information that it seems appropriate to put centrally located or to have access to it from a central place so that the agencies can use it but without sacrificing the confidentiality of individuals. We think it can be done.

Mr. GALLAGHER. Somebody decided you can do that. Just as a matter of economics, I am sure the Budget Bureau would have an interest in it.

Mr. BOWMAN. We certainly do.

Mr. GALLAGHER. How many potential customers do you have now? How many users do you have? Who will use it?

Mr. BOWMAN. I will get you information on people that are doing this now, if you wish me to. I do not have this information—

Mr. GALLAGHER. Would this not be something that in the beginning should have stimulated the desire for the Central Data Bank or is this just Mr. Ruggles' idea that we ought to have it? Is there a need for it?

Mr. BOWMAN. Let us say it is not Mr. Ruggles' idea alone. A committee was appointed because there is a clear-felt need on the part of a variety of people. They do not have now adequate access to information that is available.

Mr. GALLAGHER. Who?

Mr. BOWMAN. Various members in the academic community, various students of the economy, people who use information in order to—

Mr. GALLAGHER. Do they use it now?

Mr. BOWMAN. They now gather under various hindrances. They think this would improve their access to data without—

Mr. GALLAGHER. Do the agencies who now have the information in these tapes have a list of the users? How many requests have been made? How many requests have been denied?

Mr. BOWMAN. Yes. For some of the things in which this has been done we do have that kind of information.

Mr. GALLAGHER. Does Mr. Ruggles or Mr. Dunn or yourself have this information?

Mr. BOWMAN. I would have access to it. I have not really asked for it. How many—

Mr. GALLAGHER. Is there 10? 100? We are going to spend a lot of money.

Mr. BOWMAN. There are thousands.

Mr. GALLAGHER. How many requests have been denied? Have they ever sought information that properly was turned down? How many times?

Mr. BOWMAN. I presume they have. I cannot testify here and now in detail.

Mr. GALLAGHER. If the Budget does not know, who will know? Certainly not Mr. Ruggles or Mr. Dunn.

Mr. BOWMAN. The Budget may not. I do not know personally because I did not prepare myself for that kind of information.

Mr. GALLAGHER. The question is: We are about to depart from a long-established practice.

Mr. BOWMAN. We are not going to depart.

Mr. GALLAGHER. I would like to know whether or not there is a justification for this departure.

Mr. BOWMAN. Mr. Chairman, we are not departing from a long-established practice. We are adhering to a long-established practice. In a very clear sense any one of the large statistical agencies is a data center now.

Mr. GALLAGHER. Why do we not allow it to continue in the efficient fashion it has in the past?

Mr. BOWMAN. Because we think after we have investigated it a little more carefully that there are more efficient ways of doing so. As it is now the Federal statistical system is a decentralized one. There are many people who feel decentralization is a general advantage for the general collection and development of statistics. There are others who feel differently.

Mr. GALLAGHER. Everybody who is a statistic would agree with this?

Mr. BOWMAN. You would make it so decentralized nobody collected individual information on an individual except one agency. Then you would have an agency for every individual.

Mr. GALLAGHER. No; I do not think we should do that. I think what we are doing now is just fine. I am trying to see some justification as to why we should do it another way. Since you have not designed a system you do not know how many users there are, what the purposes will be, how much it will cost, or why we should do it, I think we ought to abandon it for a while and go back and study it.

Mr. ROSENTHAL. Mr. Chairman, if I may interject a note at this moment. You do not understand it because you have not read that well-received book, "Ruggles of Statistical Gap."

Mr. GALLAGHER. There seems to be a gap involved right here as to why we should do this, Mr. Bowman.

Mr. BOWMAN. We have not yet proposed to do it. We are considering the possibility of making such a proposal so far as this hearing is concerned. We were invited here because there is a great concern about, as there should be, various other developments that are taking place in the economy which seem to involve invasions of privacy. I am here to indicate the way in which I think this will affect a proposal on a Federal Statistical Data Center. I have tried just as clearly as I could to state my position.

I do not believe that I can improve that statement. I think the Federal Statistical Data Center would achieve, could be achieved, and not invade privacy.

Mr. GALLAGHER. You have given a great deal of study to the statistical means for this. Should there not be a corresponding study as far as the desirability of it and as far as the safeguarding of these statistics are concerned, before we start?

Mr. BOWMAN. I am glad you asked that question. It seems to me you asked that question because it seems to me I have not made one point clear.

All my comments are about a statistical data center. This center would not serve a lot of other needs for which other people would say there are needs for centers. I am not commenting on those. I cannot comment on those because they are outside my professional competence and outside of the area for which I have responsibilities. There may be many other kinds of centers proposed. There are.

Mr. GALLAGHER. Mr. Bowman, I have tried to confine our questions to your area of responsibility and to your area of competence, for which I have great respect. I have asked you several questions which I do not feel that you really have responded to. Who are the users? Why? What use has been made of the information that you now have? What compels the necessity of centralization? If you are about to centralize, what kind of a system are you suggesting?

This is one of the things that disturbs this committee.

Mr. BOWMAN. Mr. Chairman, let me answer the first one. Who are the users? At the present time the Census Bureau is a user of IRS information. Other agencies are users of IRS information. Each of them has to make their separate arrangements with IRS.

Mr. GALLAGHER. What is wrong with that?

Mr. BOWMAN. There might be considerable economy if the IRS information were available in one place, better organized for the users, to which the different Federal agencies can put it so that this could be done through one agency instead of having to make arrangements in several different ways. In other words, if the Census Bureau gets information from IRS on data tapes and pays for the cost of doing so, then the question is, if another agency wants exactly the same information and goes to IRS do they do it all over again?

Mr. GALLAGHER. I say "Yes." Let us do it all over again because you want to hand-carry an incompatible tape over to another agency.

Mr. BOWMAN. No, once it was done it would be available to all Federal agencies so long as the agency did not ask for information about individual respondents.

Mr. GALLAGHER. That is done.

Mr. BOWMAN. Did I answer that question?

Mr. GALLAGHER. Actually, you have raised the question that I really must get back to. What kind of a system is going to make this fully protective to the individual? Is there one now in existence? Is there one that has been designed to build in adequate safeguards?

Mr. BOWMAN. Yes, the same system would apply to information in the Data Center as now applies to information in the Census Bureau. They have all this information that you are talking about.

Mr. GALLAGHER. Yes, you are telling me that this can be done. Anybody in the computer business that we have spoken to advises me that it cannot be done.

Mr. BOWMAN. You are talking about techniques of communicating information.

Mr. GALLAGHER. We are either talking of hand-carrying tapes or telecommunications. What are you talking about?

Mr. BOWMAN. I am not willing to talk exclusively about either one of those but I am willing to say—

Mr. GALLAGHER. There are not many other areas involved here.

Mr. BOWMAN. If the objection is to telecommunications, I think it could be established that there are considerable advantages to a data center even if you had to hand-carry the tapes. I would not want to rule out—

Mr. GALLAGHER. Is that what you are recommending? Hand-carry tapes to agencies and use computers?

Mr. BOWMAN. Hand-carrying?

Mr. GALLAGHER. Yes. If you are ruling out telecommunications you are now—

Mr. BOWMAN. I have not ruled it out. I have said that it is not exclusively necessary in all areas.

Mr. GALLAGHER. Mr. Bowman, with all due respect, if you are ruling out telecommunications, why can you not now use the hand system we have and hand-carry tape for those people who wish access to it if they are legally entitled to access to it?

Mr. BOWMAN. Mr. Chairman, much data collected by present agencies, and which use computers, are collected in regional offices. We do not have a situation now for communicating the information in the regional offices, telecommunicationswise, in many instances to the agency which is responsible for tabulating the data. We do not have that now.

Mr. GALLAGHER. Are you recommending telecommunication now?

Mr. BOWMAN. Am I recommending it extensively in every area where it is possible?

Mr. GALLAGHER. No, for the general purposes of your central data bank and regional offices and departments that will be participants.

Mr. BOWMAN. To the extent to which I would be willing to recommend it, we have not studied it yet. You see, this hearing is in a sense preliminary to any proposal that we are bringing before the Congress. We have not finished our job yet. We have not completed our review.

Mr. CORNISH. Mr. Bowman, in one of the reports it is suggested that telecommunications might be used to provide the information from the Data Center to the users; is that true?

Mr. BOWMAN. That is right.

Mr. CORNISH. I think that possibly is one of the points that the chairman is trying to make in this regard. That is an idea which is before the Budget Bureau?

Mr. BOWMAN. That is right. I think we will want to investigate it and we will want to look very closely at whether or not this provides the opportunity for disclosing, for invading anybody's privacy. This is one of the things that will have to be examined.

Mr. GALLAGHER. It should be an essential thing that it be examined by somebody before it is set up and the same machine call off the kind of information they should not have access to.

Mr. BOWMAN. Suppose you do not reveal the identity of any individual?

Mr. GALLAGHER. Can you assure me that it would not? You are not assuring me that you will not eliminate individual identification. If you do, we can say it is a fine thing.

Mr. BOWMAN. I am not assuring you that there will be no information in the Data Center that does not reveal the identity of an individual. I have not said that I would recommend any system of telecommunications which is itself subject to revealing the identity of individuals.

Mr. HORTON. Mr. Chairman, on that point.

As a matter of fact, affirmatively, you have to have the identity of the person or the corporation in order for the statistical information to be of any benefit?

Mr. BOWMAN. In most instances; not in all.

Mr. HORTON. Let us not quibble over words. You have to, in order for this proposed Center to be effective, have the identity of the individual or the corporation furnishing this information?

Mr. BOWMAN. That is what I said in my—

Mr. HORTON. No, you did not say that.

Mr. BOWMAN. Just a minute.

Mr. HORTON. I did not understand that. Let us put it that way.

Mr. BOWMAN. I am very willing to say, and I thought I did say—and I think the transcript will show that I did say—that I could not commit myself to Mr. Gallagher's questions. I could not answer it by saying that there will be nothing in the Data Center that does not reveal—there will be nothing—the Data Center will have no information that reveals the identity of an individual.

Mr. HORTON. I do not want to quibble over words. I am saying to you that you have to state it affirmatively in order for the statistical information to be of any benefit. You have to have the identity of the individual or the corporation or the organization in that data bank.

Mr. BOWMAN. I do so state.

Mr. KRUEGER. May I comment on the question that you are raising here with regard to systems of communication. I think we can say most emphatically that unless the technological advances in the general area of telecommunications or between computers is advanced to the point where this kind of system can be used with sufficient safeguards protecting confidentiality, we would not propose its use.

Mr. HORTON. That gets back to the question of how are you going to protect that confidentiality?

Mr. KRUEGER. What we had thought of was the same kind of protection we now have which requires agencies which collect information under pledges of confidentiality to maintain that confidentiality and which have provisions in law imposing penalties for anyone who violates that.

Mr. HORTON. You are talking about a present system that is quite different from the system that you are now proposing or that you are at least considering. At the present time in many of these agencies they have computer banks already. Is this not a fact?

Mr. BOWMAN. I think we are not proposing a system that is significantly different from the system we now have. We are proposing—

Mr. HORTON. You may not feel it is that way but it seems to me that it is.



Answer this question: Are there not now in existence in various agencies throughout the Federal Government these computer banks?

Mr. BOWMAN. Yes, sir.

Mr. HORTON. What you are proposing is an octopus. You are going to put the head on all these figures because each one of these agencies has these data banks already. So what you are going to do is provide the body for these arms and you are going to put all this information in one central place and then any other department will be able to get that information, will be able to get it on a moment's notice, and conceivably the thing we are concerned about is that this type of information can become public knowledge.

Mr. BOWMAN. That is a very good way of stating it. I think to a considerable extent it represents the background of our thinking about this matter. Suppose it is true that all of these agencies that now collect statistical data, they do not all have banks of data of other agencies now but a great many of them do. Suppose they all had it. What is protecting the confidentiality of that information now?

Mr. HORTON. That is what I am concerned about.

Mr. BOWMAN. I think it would be much better. I believe it is still protected but I think it would be much better to have that information assembled in one place under strict rules of confidentiality that can be much more clearly watched than having it spread over several different agencies all of which are trying to practice rules of confidentiality but they may not all be everything that we would hope they would be.

Mr. HORTON. I understand there is a proposal—I don't know what the status of it is, but that the National Crime Center, which is on a computer basis, is being considered. They will be able to get quite a bit of information. This information, whether you hand carry it, or send it by telecommunications, will end up in this central data bank.

I notice Mr. Macy—and you are probably familiar with his article in the Saturday Review—said “for proper decisions in these areas,” meaning decisions on personnel planning, jobs, and so forth—

In forecasting manpower needs and important decisions of career planning, for proper decisions in these areas, we must have integrated information systems. This will require the use of information across departmental boundaries. It is here that current efforts to standardize symbols and codes will pay dividends. Direct tape-to-tape feeding of data from one department to another may become common.

Now, that is Mr. Macy talking. He is going to be one of the users of this information. If the President or any other agency has a demand for a certain type of personnel, they are going to make a request on your central data bank.

Mr. BOWMAN. Not our data bank, no.

It says information systems. I think this is where I agree with all the things I have heard here today. The thing we are talking about is not making or not even discussing a proposal for a single, all-purpose data bank. What I am discussing is a single-purpose data bank of not all information that is available but of information that is particularly relevant for statistical purposes which is basically quantitative information. It is measurements of one sort or another. But, irrespective of the fact that it might have some information in the statistical data center that is in another data center for other pur-

poses, so far as the statistical data center is concerned it would not release any information about the individual.

Mr. HORTON. It might not, but the other agency might.

Are you telling us now you would not make the information in your centralized data bank available to the Civil Service Commission?

Mr. BOWMAN. I am certainly doing so on an individual—I am—

Mr. HORTON. Are you saying you wouldn't make it available to the FBI?

Mr. BOWMAN. I certainly am.

Mr. HORTON. Under no circumstances would you make it available?

Mr. BOWMAN. Just as the Census Bureau would not make it available to Mr. Macy, or would not make it available to the FBI.

Mr. HORTON. But you have indicated that this information will be made available to you from the various agencies. You indicated that the Census Bureau information would be fed into this computer that you are proposing.

Mr. BOWMAN. But I am proposing that the Federal statistical data center have the same rules of confidentiality as now applies to the Census Bureau and any information in the data center will not be released by the data center in terms that allow the information to be associated with any individual.

Now, if the data center had some information in it from an agency that didn't have those restrictions, it would still be my understanding that the data center would not release that information. The agency might release it, but not the data center.

Mr. HORTON. I hate to hedge on words, but now I think we have to, and I don't like to do that, but you said the center would not release information that would furnish identification. Now, that would mean to me that you would release information.

Mr. BOWMAN. If the center had information about an individual, no matter how it got it, it was in the center, and Mr. Macy wanted that information about an individual, the center would not release it.

Mr. GALLAGHER. Who is going to control the center?

Mr. BOWMAN. Who controls the Census Bureau now? Law. The center would be controlled in the same way.

Mr. GALLAGHER. By which agency, the Census Bureau?

Mr. BOWMAN. We haven't decided where the center might be located.

Mr. GALLAGHER. Some people might pay more attention to the law than others in the Government. For instance, you give great weight to a subpoena. There are other agencies who hand out subpoenas like Kleenex, without much authority—we have had a devaluation of the subpoena process.

Now, you give great weight to it, but other agencies do not. I would be interested to know who is going to run the center.

Mr. BOWMAN. We haven't made up our minds. It might be the Census Bureau itself. It might be in the Department of Commerce closely associated with the Census Bureau. It might be in another agency.

These are the things that are now under review.

One of the considerations in deciding where it would be would be this problem of the history of maintaining confidentiality.

We, definitely, Mr. Chairman, are just as strongly in support of the idea as you are, that for the statistical program which I am mainly concerned about, we do not want in any way to violate the confidentiality of information with regard to the individual.

Mr. GALLAGHER. Assuming we are going to strictly adhere to the law, the problem of computers at this stage of the game is something that bothers us. Electronic radiation transmission problems are still in research and they are only in a research stage at this point. This is one of the reasons why at this particular moment of our time we object and would like to be convinced that confidentiality can be protected. Unless we are all going to start living in lead-lined rooms in this country, and all of the offices involved in the data bank will be lined with lead, there is the problem of anyone getting on the same transmission channel and requesting that information and acquiring that information.

Now, this just has not been solved. I don't know what your feelings are. Perhaps you think it has been solved.

Mr. BOWMAN. No, I don't think it has been solved, but I must admit I don't see its specific relevance to the main idea that we are talking about here.

Mr. GALLAGHER. Its specific relevance is very simple. If I were fortunate enough to have the computer give me a little patronage and give my uncle a job over at the data center, he might see a key lying around and therefore he could turn on the proper channel and recall information on somebody he might not particularly like, or he might have a cousin on whom he would like to do a little double-think kind of operation, and he might flash through one of the erasure-type devices and for all history that person would no longer exist.

Mr. BOWMAN. This can be done now at many of the operations.

Mr. GALLAGHER. Yes, it can. He might exist in census and might not exist in IRS, but he is going to totally not exist in a central data bank.

Mr. BOWMAN. I don't really think so, Mr. Chairman.

Mr. GALLAGHER. Well, not this week, but you know that in the interests of efficiency and economy we will have people in here 3 years from now saying we ought to put a little more information in there.

Mr. BOWMAN. You certainly realize how carefully I have tried to be—

Mr. GALLAGHER. Yes. You haven't gotten around to answering the questions I asked about an hour ago.

Mr. BOWMAN. I think that is unfair, Mr. Chairman. I think that is quite unfair.

Mr. GALLAGHER. No. No. Mr. Bowman, you have not answered my questions as to whether or not there is a system that can protect confidentiality and build in the very things upon which you have told me, and upon which I believe you, and on which we share a concurrence of opinion.

Mr. BOWMAN. If you are talking about an electronic data communications system, I don't know nearly enough to answer that question. There may not be any ways of protecting communication by that method that is completely foolproof against other people tapping in on the system. I would agree with you that the Defense Department knows a lot more about that than I do. But I would say that applies to everything that we are doing now without any changes.

I would say that so far as there being in a statistical data center the information which we now have, in the same general forms as we now have it in separate centers, I see no reason why that information in that center cannot be protected in the same way as it is being protected in the 20 or 25 centers in which we now have it.

Mr. GALLAGHER. You don't see a shade of difference in having that information protected by the law, but subject to a tampering with the law in 25 sections and putting all of this into one area where the same weaknesses could exist?

Mr. BOWMAN. If I see a difference, I see an answer in the centralization.

Mr. GALLAGHER. You don't see the disadvantages to the individual's problem, or the confidentiality aspects? You see no difference in having a centrally located center of information and having it presently exist in 20 or 25 different areas?

The risk has been reduced to acquire information.

Mr. BOWMAN. I don't believe putting it in various pools has significantly reduced the risk, no.

Mr. GALLAGHER. If there is a breakdown in the security of one agency, there is not necessarily a breakdown in the other agencies, but if there is a breakdown in the central data bank, it seems to me we now have a problem.

Mr. BOWMAN. There is, of course, the problem of whether you have all your eggs in one basket or in different baskets, but I would say a well regulated data—

Mr. GALLAGHER. Right there, the egg in the basket. Supposing someone picks up the one basket with all the eggs?

Mr. BOWMAN. You get them all broken. On the other hand, suppose you are very careful about handling the big basket with all the eggs in it and very careless about handling the—

Mr. GALLAGHER. You have not assured me that we have an egg carrier or an egg transmission system fully capable of doing this.

Mr. BOWMAN. Not a transmission system, but I hope I have assured you, Mr. Chairman, that if you can develop a practice of confidentiality as we have in the statistical program, if you can develop that practice and the whole business community and everybody else has confidence in it, that we could develop an agency—maybe part of one of these agencies—that would have all of the data of the various statistical agencies for general use, for which this same reputation would be present.

I hope I have assured you of that. Maybe I haven't.

Mr. GALLAGHER. You have assured me, Mr. Bowman, of your very good intentions. I am not assured that those good intentions can be carried out. That is what worries us. That is what worried us at the very beginning.

May I ask you just one question? Do you now have an inventory of statistical information that is centrally available?

Mr. BOWMAN. An inventory of all statistical information?

Mr. GALLAGHER. That all of the agencies now have.

Mr. BOWMAN. No.

Mr. GALLAGHER. Wouldn't it be cheaper or more efficient if you had an inventory of what data is available in the various agencies now and after a period of a year or two, add up how many requests have

been made for the telephone, how many requests have been denied, and what the problems are, before we get down into a whole new area? It would seem to me that would be more efficient, far more cheaper and protective of the individual. If we haven't done this now, why should we now move into a central data bank when the need has not yet been created or determined?

Mr. BOWMAN. At the back of the Ruggles report there is a practical inventory we made intimating some of the things which seem to be in a form now which would be readily available without us making a decision that we would want to put all of them in the data center, but which seem to be readily available for this purpose. This gives you some idea of some of the things that could be put into a data center. We—

Mr. GALLAGHER. I have read that, Mr. Bowman. What I am asking is, has there ever been an inventory published of the kind of data that is now available—even though it is not centralized.

Mr. BOWMAN. That is the closest thing that I know of to that sort of thing.

Mr. GALLAGHER. Would it not be more advisable to have a full inventory and then distribute it to the potential users to see how many people or what kind of users might want it and how reasonable their requests are before we got into what appears to be the start of a whole new area in government, and not a very efficient operation?

Mr. BOWMAN. We communicate with our public now to a great extent through publications. The inadequacies of publications for many purposes are the things that Mr. Ruggles and the other people of that sort have been talking about.

It would be our hope if we had a statistical data center that operated so that confidentiality could be maintained in the way you say my intentions are, that some of the detailed publications that we now make could be avoided. We might make some economies along those lines. We can never serve the real needs by these detailed publications.

You see, we are all caught in a new system of technology here. We are trying to think our way through it.

Mr. GALLAGHER. We are trying to make sure the people are not caught in the new system. That is why we are here today.

Mr. BOWMAN. Certainly on that basis I am 100 percent in favor of the work of this committee.

Mr. GALLAGHER. Would you consider gathering information and publishing an inventory, before we got into the central data bank, to see what the demands are for this kind of information?

Mr. KRUEGER. It would be the interest in this kind of thing which led the so-called Ruggles Committee to study the whole problem and led them to make a report to the Social Science Research Council in the first instance. These were essentially groups of people—there were people also in the Government concerned with the fact that there is available now a great deal of information; it is stored around in different places; in order to make the kind of uses for statistical analysis, they would like to—they find it difficult to get access to it under present administrative arrangements.

Mr. GALLAGHER. That is not all bad.

Mr. KRUEGER. Not all bad, no, but part of it is and it is the bad part with which we are concerned, which we conceived of, as the formation of this kind of capability as providing a service for.

Mr. GALLAGHER. May I rephrase the question?

Is there now a complete inventory as to the kind of statistics that are now available? For instance, in the 9,000 tapes that you mentioned in your report.

Mr. KRUEGER. No, not a complete inventory.

Mr. GALLAGHER. Has there been a request of Mr. Ruggles and his organization and others who might be interested in this type of information, to make such an inventory and make such a publication to see how many people might want access to this information?

Mr. KRUEGER. We made enough of an inventory to come to the conclusion there is a sufficient amount of very useful information in which a sufficient number of people both inside and outside the government would have interest in, to think that we were warranted in pursuing the idea further and that is essentially where we are now.

Mr. GALLAGHER. How many users make a call on this information?

Mr. KRUEGER. I don't know. That in itself would be a very difficult statistic to get.

Mr. GALLAGHER. This is something we ought to computerize. Have there been 100 people who would want this kind of information? A thousand? How many people have been turned down for making an unreasonable request?

Before we start spending millions of dollars, it would seem to me we ought to determine how many potential users there are. It may well be there are just one or two organizations who are interested in this and perhaps the taxpayer shouldn't be asked to assume this particular burden.

Mr. BOWMAN. Our evaluation will try to take this into account before we come to a regular proposal.

Mr. GALLAGHER. Thank you.

Mr. Rosenthal.

Mr. ROSENTHAL. Thank you, Mr. Chairman.

How much money have you spent to date in all studies and surveys leading to the point where we are now?

Mr. BOWMAN. The Ruggles Committee didn't cost us anything.

Mr. KRUEGER. The Dunn report cost a little over \$12,000.

Mr. ROSENTHAL. At the moment that is the full amount that has been expended in this endeavor?

Mr. KRUEGER. Yes.

Mr. ROSENTHAL. At the moment you really don't have any idea how many users you would have or how many agencies would want it? You are going along on the general assumption that such a thing is useful?

Mr. KRUEGER. Yes.

Mr. ROSENTHAL. Mr. Bowman, is that your position?

Mr. BOWMAN. We are going along on the assumption there has been demonstrated a significant demand for this sort of thing both within the Federal agencies and among the general public but we don't have an actual count of the number of users.

Mr. ROSENTHAL. Has anyone directed a specific inquiry to you asking that such a center be set up?

Mr. BOWMAN. Has any particular agency asked for it? Part of the work Mr. Dunn did was to talk with the agencies. I don't know that there was any particular request for setting up a center.

Mr. ROSENTHAL. Is the movement for this from the inside out or the outside in? Is this a part of a small group who thought this would be a good idea and went out and solicited potential customers, or did the agencies themselves find a need for this and suggest it to the centralized headquarters?

Mr. BOWMAN. Let me answer the question. I have been with the Bureau of the Budget now for 10 years. This has been discussed and suggested in innumerable conversations with me by almost every agency in the Federal Government, yes, and by all sorts of persons outside the Federal Government, State and local governments, individual scholars and a great many people, that there is need for better access and better ability to use the data which is now distributed among different agencies.

Mr. ROSENTHAL. Have you found that private industry has been in the forefront of making use of computers in modern technology?

Mr. BOWMAN. Large industry has, yes.

Mr. ROSENTHAL. Is it part of your idea to have the Government keep up with private industry?

Mr. BOWMAN. I would hope the Government would always be as efficient as possible with modern technology, yes.

Mr. ROSENTHAL. In the case of computerized equipment has private industry taken the lead as compared to the Government?

Mr. BOWMAN. I don't believe so.

Mr. ROSENTHAL. You would agree with the proposition that the Government has special responsibilities in terms of constitutional rights and invasion of privacy that private industry might not have?

Mr. BOWMAN. I certainly do.

Mr. ROSENTHAL. And that in your use of technological equipment, there is this unusual and much harsher but appropriate burden of responding to constitutional obligations?

Mr. BOWMAN. And I would agree that in some instances modern technology might have to be not pressed as far as it could go in order to protect the constitutional rights of individuals.

Mr. ROSENTHAL. Subsequent to this meeting and perhaps before that, you had that as a principle, that you might have to give up some of the advantages of modern technology in terms of the constitutional responsibility?

Mr. BOWMAN. That is right.

Mr. ROSENTHAL. I note nowhere in the Dunn report any comment at all that we have to be aware of invasion of privacy, or individual privacy, or constitutional responsibilities. The closest I come to it is on page 10, the disclosure problem. In very few words nothing is mentioned at all in the whole report about this area of special Federal responsibility.

Mr. BOWMAN. The only way I can explain this, Mr. Rosenthal, is that all of us who were working in this area sort of assumed that it didn't really have to be discussed. In other words, the statistical system has been so imbued with the notion that you do not reveal information about the individual—it didn't get mentioned. I am very sorry, and I think Mr. Dunn is probably very sorry, but you can ask him.

Mr. ROSENTHAL. But those in the statistical field in the Federal Government have a special role as compared to those in private industry, in this area of constitutional rights and constitutional obligations.

Mr. BOWMAN. Yes. I think we have a special role. It is not, however, unknown in business. In other words, banks have a recognition of the privileged character of your deposits and other information they may have about you. Doctors have it. Lawyers have it.

Mr. ROSENTHAL. On the contrary, in terms of credit information, I think banks and commercial institutions exchange information quite freely and I think they consider that is one of their commercial assets.

In your statement, Mr. Bowman, you list the members of the task force. They are all very distinguished men. However, I do not note that any of them have made their special mark historically in areas of constitutional responsibilities or awareness of invasions of privacy.

In other words, they are all the statistical type, would you say that?

Mr. BOWMAN. Yes.

Mr. ROSENTHAL. In other words, what bothers me is that—I think the same thing applies to all the members of the committee—there is a group within Government who are statistically oriented and are desirous of following the lead of private industry and taking advantage of computerized facilities who are pushing ahead rather vigorously and yet have taken no overt step either to acknowledge to themselves, to the Congress or to the public, that they have a special responsibility in the field of invasion of privacy.

Mr. BOWMAN. But I haven't done that, have I?

Mr. ROSENTHAL. Until today, what have you done overtly to indicate that you considered this a special responsibility?

Mr. BOWMAN. I did illustrate in my testimony the office and the Bureau of the Budget were very active. The *St. Regis* case; we were active through the lower courts and up to the Supreme Court. We were active in the position that the Solicitor General took before the Supreme Court, which was to argue against the right of the Federal Trade Commission to get the copy of the return from the *St. Regis* Co. The Supreme Court decided elsewhere.

We were active also in helping to sponsor the legislation that eventually corrected this.

All I can say is that the activities of the Bureau of the Budget, my office in particular, have been continuously in the direction of maintaining the confidentiality of statistical information, against its being transferred to agencies that would use it largely for purposes of prosecution or actions against individuals. So the Federal Trade Commission knows that if they came to us with the idea of getting census information that we would strongly support the Census Bureau. If they came to us with the idea of getting information from another statistical agency, we strongly support the nontransfer of that information which identified any individual.

Now, if you say in the writing of these various reports in this area, were we lax in not paying more attention to the writeup of this particular problem, if you want me to plead guilty to that I will plead guilty. Yes, we probably should have made some more mention.

Mr. ROSENTHAL. I am not worried about the writeup. That is a trivial thing as far as I am concerned. I am worried about a commitment—as to the nature of your commitment. I have a sneaking suspicion your commitment lies in the area of getting real sophisticated automated technology into the Federal Government so that you can keep abreast of what the large corporations are doing.

I would like to see that commitment matched with the special kind of commitment that we in Congress feel we have to the people of preserving constitutional rights.

The difference of opinion between you and the committee, as I see it, is that we are both in favor of advancing the use of technological equipment.

You think this can be done and still preserve the inherent rights that people have under the Constitution. Frankly, from the testimony here this morning and from yesterday's testimony and other things, I have grave doubts and deep reservations that the national data center can be developed without infringing on individuals constitutional rights—unless Congress enacts laws to prevent you from doing the things that we think should be done.

We can't rely on your good judgment or the good judgment of your associates to prevent these things from happening. Before this center is developed, before any money is authorized or appropriated, I think we are going to have to enact laws to make sure that there are no invasions and that the privileges you are permitted to have to make use of this machine are not abused. This is where our difference of opinion is.

You think we can rely on your good judgment. My own feeling is that we are going to have to rely on legislation.

Mr. BOWMAN. I don't really disagree with what you have said and it may well be that the statistical data center, if it is set up, will require and need, and we will propose that it specify, the passage of laws to protect the confidentiality of information that is stored within it. I don't disagree with you on that.

If I disagree at all, it is merely a matter of saying I don't really want to talk about data centers broadly. A statistical data center where the information is to be used for statistical purposes, I think law can be devised which will protect the confidentiality of information so far as the individual is concerned.

I would also like to answer this question. I think I am equally, as strongly convinced as anyone, that the constitutional protection of individuals is just as uppermost in my mind as is the economy of operations of the Government.

On that point, I am a strong supporter of this point.

Now, with regard to other types of data centers, it may well be that the law will have to make up its mind whether it is desirable or undesirable to have information centralized of a certain kind.

For example, we now have a center for the registration of automobile operation registration. The States and the Federal Government operate this center so that one State can find out whether or not a revocation has taken place in one State that the other State doesn't know about.

Now, this is definitely an infringement, if you will, of some kind on the individual.

Mr. ROSENTHAL. That is not quite the same as sending an employment letter to a professor in school and saying, "What do you think about this fellow? What is his temperament and personality?"

If a man's license is revoked, I see no reason not to tell every State about it.

Mr. BOWMAN. That is the way I feel about it, but every now and again I have a feeling that people are including this as well.

For instance, the writing of a letter to a professor and asking him what he thinks about a person, I can't possibly conceive how that would even get in our Federal statistical data center.

Mr. ROSENTHAL. It would get into an employment or personnel file in an agency and I can conceive that that file in some way could be transmitted or the information therein sent to a Federal data center.

Mr. BOWMAN. A Federal center, but I hope not our Federal statistical data center.

Mr. ROSENTHAL. I don't know a thing about the statistical center and I can visualize this kind of information being gathered and kept in such a center.

Mr. BOWMAN. If it did get there, it shouldn't get out as far as the individual is concerned.

Mr. ROSENTHAL. Right.

Mr. BOWMAN. There will have to be legislation which I hope you will submit and if you fail to act in that regard I suppose the members of this committee will have to prepare such legislation prior to and as a condition, in my judgment, a condition to the establishment of such a concern.

Mr. CORNISH. Mr. Bowman, we understand your testimony is based on the rather current proposals that are now before the Bureau of the Budget, but I hope you will understand that our concern extends into the future a number of years where there may be many pressures and demands to make this computer center something more than what it is presently proposed as.

In that context, I hope you will understand the concern of the committee.

Mr. ROSENTHAL. Thank you, Mr. Chairman.

Mr. GALLAGHER. This might well be incidental, yet it points out the problem: This scientific data business that is going on is all well intentioned, where boys and girls write into a data center and give a lot of information as to the size of their moles and all sorts of other things. That was a lot of fun at the beginning, but now people are running off and selling tapes, catering to the kinds of feelings that they might have.

There is a basic weakness in this also. I think one of the advertisements advertised a \$2.50 a year membership, or a \$400 life membership, which is for the real losers. It shows that the computer really does not solve all the problems. It is one of the things that we worry about, that it might create more problems than it started out solving, and create new problems. Now it is suggested that millions of people will be programed and the lists might be sold for other purposes. So what starts out well intentioned actually could end up very seriously an invasion of people's privacy. All of the bugs have not been eliminated.

Let me ask you this: Is this presently funded, now? Could you go ahead now?

Mr. BOWMAN. There would have to be an appropriation.

Mr. GALLAGHER. But you have a contingency fund.

Mr. BOWMAN. The contingency fund is merely an allowance in the budget but the funds that would come out of the contingency fund would still have to be appropriated. Is that right?



Mr. KRUEGER. It is a contingency allowance in the budget. There is no fund in the sense that here is a pot of money somebody can reach into and get anything out of.

Mr. GALLAGHER. Is it funded now? Can you start it now without any further legislation?

Mr. CORNISH. Using the suggested 9,000 tapes.

Mr. GALLAGHER. The 9,000 tapes that Dr. Dunn mentioned—could you now start centralizing them with the existing contingency fund that you have?

Mr. KRUEGER. I would guess probably that any agency which had funds not required for other uses would have a little difficulty explaining how they might use the funds to go off on this kind of an enterprise without some kind of authorization.

Mr. GALLAGHER. I notice there is a contingency fund in the U.S. budget.

Mr. KRUEGER. If you are talking about the contingency allowance in the budget, that is \$500 million.

Mr. GALLAGHER. Would \$400,000 of that \$500 million have some relationship to the Data Center?

Mr. KRUEGER. Not that I know of.

Mr. ROSENTHAL. You are not going to go ahead and do this without congressional approval, are you?

Mr. GALLAGHER. This is what we want to know here today.

Mr. BOWMAN. No.

Mr. ROSENTHAL. The answer is "No"?

Mr. BOWMAN. Yes.

Mr. ROSENTHAL. Thank you.

Mr. GALLAGHER. You will come up with an answer and a request before this thing gets underway?

Mr. BOWMAN. At the present time we are merely considering this. It probably would be possible for an agency to take on some of the functions of the Data Center, but we have no intention of doing this without a proposal that would be officially presented to the Congress.

Mr. GALLAGHER. And we have now your assurance as part of a legislative record which we are trying to establish here, that when and if your studies are concluded, before you will go ahead, you will come to the Congress and request permission, is that correct?

Mr. BOWMAN. That is right. We would propose it probably as a regular budgeted item and if it requires any special legislation the legislation would have to be drawn and it might well require special legislation with regard to confidentiality.

Mr. GALLAGHER. You might want to propose that when the Kaysen committee report is finished, that it be submitted to a distinguished panel of experts on constitutional law, and invasion of privacy, for their comment and suggestion.

This review would help spotlight the problems that perhaps the statistical people might not consider in their proposals.

Would you feel that that might be a helpful way of bringing about this proposal or helpful in your final decision?

Mr. BOWMAN. Certainly I will discuss this with the Director who asked for this report and indicate this is one of the things that you suggested.

Mr. GALLAGHER. We have two candidates named Reich and Packard whom we would like to recommend for sitting in on such a panel.

As I say, this would be helpful. I think what we have to do is to make certain that we are not going to adopt this without having a real hard look at it from every aspect, from the law, from the invasion of privacy, from the interests of the citizen, as opposed to just the statistical needs.

Now, I would hope that you could consider that suggestion.

Mr. BOWMAN. We shall do so.

Mr. GALLAGHER. Mr. Horton.

Mr. HORTON. Mr. Chairman, thank you.

I start off with the premise that I am not mad at computers and that computers can do good things for us and in my opening statement yesterday I tried to indicate a computer can be very helpful if it is utilized in a proper way.

My approach to this is to try to make it possible for the Federal Government to use computers, but at the same time try to recognize the problem of protecting the individual.

Now, I am a bit confused because I understood that your proposal, or at least the thing that was being thought about, was not so limited as your testimony would indicate here today.

You have tended to indicate that the information that you are going to gather is statistical information and that this would just be a furnishing of statistics to other Government agencies and other users.

Is that accurate? Have you limited the proposal here today? It is not going to be a central data bank for the entire Government; is that right?

Mr. BOWMAN. That is right. Covering all functions.

Mr. HORTON. But you are going to have to get a lot of information. Where are you going to get this information?

Mr. BOWMAN. The idea here was quite broad, that all information useful for statistical purposes and to be used only for statistical purposes would go into the data center.

Mr. HORTON. I don't expect you to have now a definition of "information for statistical purposes," but I wish you would submit to this committee a definition of what you understand "information for statistical purposes" to be. I would like to know that in very much detail.

Mr. BOWMAN. Let me say what I have stated the other way around. It definitely does not involve giving any information that can be identified with any individual. It excludes that. I will try to write out a definition of what it includes.

Mr. HORTON. The present concept excludes giving out information that would be identifiable insofar as an individual corporation is concerned?

Mr. BOWMAN. Yes.

Mr. HORTON. But you have indicated earlier that the information you obtain would identify an individual and would identify a corporation or an organization—would identify the source of information, in other words, and would be very specific.

Mr. BOWMAN. That is right.

Mr. HORTON. So it would be retrievable, would it not, on the data bank?

Mr. BOWMAN. Yes.

Mr. HORTON. What you are proposing also is a bank, a computer bank that will gather all information from all the Federal agencies for statistical purposes only. Is that right?

Mr. BOWMAN. It would gather information from Federal agencies that would be used for statistical purposes, yes. It might include data collected by an agency that didn't collect it for statistical purposes originally. The Internal Revenue—

Mr. HORTON. Would you ask Mr. Macy's department, the Civil Service Commission, for information that they might have on their data bank?

Mr. BOWMAN. I wouldn't think that we would, but I am not sure that there might not be some information that he had that could be useful in a statistical data bank.

Mr. HORTON. Knowing how many people applied for a job at some point, wouldn't that be a helpful statistic?

Mr. BOWMAN. That might be.

Mr. HORTON. Or whether they were male or female?

Mr. BOWMAN. Yes.

Mr. HORTON. Whether they were black or white?

Mr. BOWMAN. Yes.

Mr. HORTON. Whether they were divorced or not divorced?

Mr. BOWMAN. The latter I don't know. I think we might have all the information that he has that is in a form that—

Mr. HORTON. We are talking about statistics now, and not individuals. This would be an interesting statistic or a helpful statistic to somebody, wouldn't it?

Mr. BOWMAN. The thing that I would not see as going into a data center is a written descriptive record about—

Mr. HORTON. I am not talking about any written descriptive record of anybody, but I am asking you if you might not have a call on the data bank of the Civil Service Commission with regard to all the information they had about individuals in a certain category?

Mr. BOWMAN. And the answer is that we might. We might want to tap that into the data center and we might also want to have actual tapes of this information that they have physically in the data center.

Mr. HORTON. Now, if you got that tape and put it in your data center, then it would be retrievable, would it not?

Mr. BOWMAN. It would be, but no one would get it from the data center.

Mr. HORTON. I am not talking about now.

Mr. BOWMAN. It would be retrievable.

Mr. HORTON. I understand your real concern about invasion of the right of privacy at this point, but I am now just asking about the information that could be accumulated and you could accumulate this information.

Mr. BOWMAN. That is right.

Mr. HORTON. In all probability you would accumulate that type of information?

Mr. BOWMAN. In all probability we would accumulate that type of information. We probably would not assemble it by individuals.

Mr. HORTON. All right. Now, if the FBI has a computerized system in which they have all the crime information in the country, might you not want to get that information also?

Mr. BOWMAN. If it were the same kind that you ascribed to the Civil Service Commission, yes, but again we would not want the records with regard to individual criminals, which are written records.

Mr. HORTON. Wouldn't you want to know how many burglars were convicted in a given year, how many people committed murder and all this sort of information?

Mr. BOWMAN. Yes.

Mr. HORTON. You would have to get the specific information from the FBI computer, would you not?

Mr. BOWMAN. That is right.

Mr. HORTON. Have you made a survey as to what agencies have these computers at the present time?

Mr. BOWMAN. The Bureau of the Budget does have a survey of all of the different agencies and this information is available. We haven't a complete inventory of all of the materials of a statistical character that are available in all of the different agencies.

Mr. HORTON. But you have the information with regard to which agencies have these computerized systems?

Mr. BOWMAN. Yes.

Mr. HORTON. Do you know which ones have it now?

Mr. BOWMAN. I cannot name them.

Mr. HORTON. You can furnish that for the record.

Mr. BOWMAN. There is a publication of the Bureau of the Budget that furnishes that.

Mr. HORTON. I think it would be helpful to have this in the record, Mr. Chairman, and I ask that we have this information.

Mr. GALLAGHER. Yes. Without objection, it is so ordered.

(The information appears in app. 1, p. 195.)

Mr. HORTON. Have you any concept of the amount of money the Federal Government has invested in these computer banks at the present time?

Mr. KRUEGER. It is in the same report.

Mr. HORTON. Do you know what it is?

Mr. KRUEGER. Offhand, I do not.

Mr. BOWMAN. It involves the statistical agencies.

Mr. HORTON. I could go through different other agencies that might have these particular banks and you would have this information. You got into some discussion with the chairman as to whether they would be hand-carried or sent over by teletype, and all this type of information. You would have to get some of this information from these various agencies. Would you not have to have the law changed in order to get this information from these agencies which would be so important to you?

Mr. BOWMAN. We might have to have it changed for some agencies. It might, however, be just the same as now—agencies can transfer information to Archives where it is stored. We have not really investigated exactly what limitations there would be with regard to the transfer of information from agencies that now have it into the data center.

Mr. HORTON. Is there not a law which says you cannot transmit information from one agency to another except in the form of statistical totals or summaries?

Mr. BOWMAN. It does not say that you cannot transmit it. It says when you do take it, you cannot release it without—

Mr. HORTON. Reading from section 139(b) of title V of the United States Code:

Information obtained by a Federal agency from any person or persons may, pursuant to section 139-139f of this title, be released to any other Federal agency only if (1) the information shall be released in the form of statistical totals or summaries; or (2) the information as supplied by persons to a Federal agency shall not, at the time of collection, have been declared by that agency or by any superior authority to be confidential; or (3) the persons supplying the information shall consent to the release of it to a second agency by the agency to which the information was originally supplied; or (4) the Federal agency to which another Federal agency shall release the information has authority to collect the information itself and such authority is supported by legal provision for criminal penalties against persons failing to supply such information.

So, you would have to have some type of authority to get this.

Mr. BOWMAN. In many instances, that is right.

Mr. HORTON. Have you talked with any of these various collecting agencies with regard to this proposal so you could find out what the problems are with regard to setting up a central data bank?

Mr. BOWMAN. We have talked with them about some of the problems that are associated with it, but we have not in detail gone into what the limitations would be on each agency. It would be more severe on some agencies than it is on others, because of the pledges of confidentiality that they have.

Mr. KRUEGER. I can illustrate how that might work by an example of what goes on. The Census Bureau operates under very strict statutory limitations which provide that information reports submitted to the Census Bureau can be examined only by sworn employees of the Department of Commerce.

Mr. ROSENTHAL. What kind of employees?

Mr. KRUEGER. Sworn employees of the Department of Commerce. The Internal Revenue collects, as you know, income tax returns. They are collected under certain other kinds of restrictions as to the use, availability, et cetera. The Census Bureau and the Internal Revenue Service have been cooperating in an undertaking which involves the collation of information obtained from the economic censuses of 1963 with information from tax returns.

Mr. HORTON. So you are saying there already is an exchange of information between these two agencies.

Mr. KRUEGER. Yes.

Mr. BOWMAN. One-way exchange.

Mr. KRUEGER. Internal Revenue, because of the particular restrictions, is able to make the tax return information available to the Census Bureau, but the Census Bureau cannot make the information from the census of manufactures, say, available to Internal Revenue.

Mr. HORTON. This is a loophole we will have to study, because one of the things I am concerned about is that this thing has been growing up helter-skelter, and nobody has been looking at it specifically as we have.

One of the witnesses yesterday pointed out with regard to the use of the social security number how it began as a confidential piece of information and now it is available to everybody, practically. So, there has been a movement away from the original intention.

I want to underscore something I think we ought to remember all the way through these hearings because I think it sums it up better than any of us can; that is what Mr. Packard said yesterday. He said:

My own hunch is that Big Brother, if he ever comes to these United States, may turn out to be not a greedy power seeker but rather a relentless bureaucrat obsessed with efficiency. And he, more than the simple power seeker, could lead us to that ultimate of horrors: a humanity in chains of plastic tape.

It seems to me what you are doing here for the sake of efficiency is to propose a source of information which can very well get out of hand because you will have to have, for your purposes, identification on that machine.

I give you the benefit of good intentions at this point, but you may not be here 5 years from now and you may not be here 10 years from now. All this information is going to be in that central bank. Somebody, for the sake of efficiency, will say, "We ought to have it for some other purpose." I cannot even envision what those purposes may be. We have to be concerned—and I hope you will be concerned particularly now—about how we are to protect that individual from having that information disseminated on an individual basis. I do not think we have sufficient safeguards at the present time.

Mr. BOWMAN. It is certainly one of my concerns.

Mr. HORTON. Do the people in your Bureau take a close look at what is already being done in these computer centers in the other agencies?

Mr. BOWMAN. Within the statistical agencies, we are quite familiar. We are not familiar with the uses of data in the nonstatistical agencies. For instance, we are not familiar with the uses of data in the Defense Department. We are not familiar with the uses of data in the FBI, the fingerprint files. So far as the use of data in the Census Bureau, the Bureau of Labor Statistics, the National Center for Health Statistics, and all of the areas of our particular concern, we are familiar with the practices which they follow in the use of the computers and the effect it has upon confidentiality.

Mr. HORTON. Do you know of any of these agencies that have come to Congress and gotten the specific authority of Congress to proceed with these so-called data banks?

Mr. BOWMAN. There are some recent acts that were passed. They were data banks broader than statistical data banks. I do not believe any statistical agencies have had any particular authorization in this area. There have been appropriations for their setting up computer facilities.

Mr. HORTON. These computer facilities are more than just statistical, because they have personal information in them. I do not understand what Mr. Macy is talking about, and I wish you would explain it to me, because Mr. Macy indicated there are already integrated information systems, and that they use this information across departmental boundaries. He says it is going to be common to use direct tape-to-tape feeding of data from one department to another. He indicates in this article that they have already used this system to give the President some recommendations with regard to filling high-level jobs.

You would have all that information on that tape. He has only limited information. I assume he does not have what the FBI has on theirs.

Mr. Macy is a very close personal friend, and I have the highest regard for him and I know he is concerned about this problem, but will it not be a lot more efficient 5 years from now for whoever is the Chairman of the Civil Service Commission to say, "We are going to need the information from the central data bank and you have it now in the Bureau of the Budget," and make a request of you and somebody is going to have to decide whether they release it, and they will find some way to release it because it will save the taxpayers a lot of money?

That is a hypothetical question, but it could happen, could it not?

Mr. BOWMAN. It could happen. The thing is, I would think much of the information would not be in the type of data center I am talking about. I did answer all your questions by saying the Civil Service Commission has information of a quantitative character about the number of people, their length of service in the Federal Government, their color, their race, their marital status, and it might well be in the data center.

I also said they could not get it back again if they wanted it so far as an individual is concerned, or no one else could, other than the Civil Service Commission.

Now with regard to files that various agencies may have, with regard to the whole history of a person's employment in the Federal Government, I can see how the Federal Government in order to operate has to know about Mr. X when he worked in this agency or that agency or another agency, and that this information can conveniently be collected and made more accessible. I am not saying that this is inappropriate. I am saying, however, it has risks associated with it that the data center which I am discussing does not seem to have, in my opinion.

Mr. HORTON. I am pointing out the possibilities and I am also thinking in terms of the fact that there are applications by law students, as was demonstrated yesterday, for employment in the Federal Government, and the Federal Government sends out questionnaires to these law professors and they put personal opinions on those reports. That information can very easily be fed into a computer.

I am concerned about the transmittal and the error in transmittal, too, because this can happen. I had a lawsuit one time where a fellow was charged with a hit-and-run accident in New York City, and he had a pretty bad situation facing him because they had his license. The sergeant had made an error and put the wrong license number down. So, this client of mine was having a really serious problem.

That information will be fed in and it would be put on a tape in Mr. Macy's agency or the FBI or the State Department or wherever it might happen to be. Conceivably, in 5 years from now we will not have a Secretary of State, so somebody will come along and want to get the information on all the people who have studied constitutional law and all the other things that he thinks go into making a good Secretary of State. Then he will go to the most efficient way of getting it, the central data bank. You will be furnishing personal information on somebody which can be reviewed and seen, and it will have a marked effect upon that person's life.

Mr. BOWMAN. We could not release it. I think you have made another point which I think is very important. It may be that we will have to consider not only regulations with regard to what comes out of the data center that I have been talking about, but also regulations about what goes into it.

Mr. HORTON. That is the point.

Mr. BOWMAN. I am perfectly amenable to this type of suggestion.

Mr. HORTON. What goes into that data bank will be very important. I do not think your study or your committee or your task force has been doing much in that connection. You will have a hard time limiting it, because you will be seeking this information from agencies that have it. It will be a lot more efficient for you to get this information from the Census Bureau, and there are some personal questions that are asked by the Census Bureau. There will be a request sometime soon to get that information.

The point I am making is that you have to guard against that.

One other point I want to make is this: Recently the Congress enacted a Freedom of Information Act. Have you looked at that in connection with the implications which might be involved?

Mr. BOWMAN. Yes, I have. In fact, I looked at it very carefully in connection with my business associates, and have been reasonably well convinced that it does not interfere with the confidentiality practices so far as the statistical program is concerned. It does have an exemption in it which indicates that where information has been collected when confidentiality is pledged, it is protected against making the information available.

Mr. HORTON. There is another point I want to make. In your statement you made reference to the fact that there is an exchange of information on Federal tax data to the Census Bureau which is relieving about 1 million small businessmen. I serve on the Small Business Committee of the House, and I am very much concerned about the problems of the small businessman.

One of the complaints that they have is the complaint about so much paperwork. I can assure you from my experience and my personal contact with a lot of these small businessmen, they do not know that this information is being shared. I think this is a problem of communication to the individual that this information is being used for one purpose or another purpose.

In other words, when he submits income tax information, he ought to know that this information is going to be made available to the Census Bureau or this information is going to be made available to the central data bank, and that fact ought to be on the form that he fills out. It is not on the form now, so far as I know. He does not know this. So, when he furnishes this information, one of the problems that I think we are concerned about is how does he get a chance to know what is in that bank so he can contest it if he wishes, or so his rights, whatever they might happen to be, can be represented.

I do not think this point has been considered. Perhaps you might give some consideration to this before you come back with a proposal to form an octopus that is going to gather all this information so all you have to do is push a key and find out all you want about a particular individual, even if you have all these safeguards that you have proposed.

Mr. BOWMAN. I think it is true a great many people do not know, and the only thing I can say is that the transfer of information from the Internal Revenue to the Census Bureau is under provisions of law as it exists.

Mr. HORTON. The law says we are all presumed to know what the law is, but the people do not know what the law is. I have one right

now under the Civil Rights Act. The New York State law is a very wide law permitting people to purchase housing. I have just had a very personal experience. I would say 99 out of 100 people do not know what that New York law is.

That is true of the Federal law, too.

I think when you are getting into this field of personal rights, individual rights and liberty, if you will, they should be put on direct notice that this information will be used in a computer and that type of information will be compiled on them. They ought to have that information in the first instance. I hope your committee will take a look at that.

You see, your task force mission, according to your own statement, was to consider measures which should be taken to improve the "storage of" and then you say "and access to U.S. Government statistics." You put "statistics" in there, but you are talking about the "storage of." You want to improve the storage of and you want to improve the access to this information.

You are building quite a monster here, and I think you will have to watch it pretty carefully before you start centralizing all this type of information that you have.

Incidentally, I am concerned about what has already happened in this field—the storage of information in the FBI, in Mr. Macy's agency, in the Census Bureau, and in the Internal Revenue Service. I think it is a good thing this committee is taking a look at this so we can awaken some people to the problems of what they are facing with regard to giving this type of information.

Thank you, Mr. Chairman.

Mr. GALLAGHER. Thank you, Mr. Horton.

I share Mr. Horton's concern. If this information is available, there will be great temptation to acquire it. You can give us all the assurances in the world that not very many people are going to have access to it, but I just wonder, if no one is going to be interested in it, whether we might be building a great and expensive electronic "garbage pail."

Mr. CORNISH. Mr. Bowman, as it now stands, each individual agency which has this material on tape is responsible for the confidentiality of that material. Is that not true?

Mr. BOWMAN. Yes.

Mr. CORNISH. So, if we set up a national data center, we would impose upon them the same confidentiality restraints, that now exist on the separate agencies. Is that correct?

Mr. BOWMAN. I think more severe than exists on some separate agencies whose data would be in the Center.

Mr. CORNISH. My point is, for example, IRS now has the responsibility of guarding the confidentiality of the tax returns. This proposal actually would introduce a second agency into that same responsibility. So, not only would IRS have the responsibility of guarding the original data that are collected, but also the Data Center would have the same responsibility for any of the material which it got from IRS. So there would be two agencies that would have responsibility for guarding confidentiality, whereas now there is only one in each case.

Do you see any problem at all in the fact that we are spreading the control out a little farther?

Mr. BOWMAN. I think that is a very excellent question. The way I tried to write my testimony and the way I have seen it to date is that the restriction in the Center would be the severest restriction. In other words, continuing the illustration you used, IRS can now make information available to States about individuals. The Data Center could not, under the way I have set it up.

That does not mean that the data could not now be made available to States. It could, because IRS would still have the right to make the data available. They would still have the data.

Mr. GALLAGHER. Are there not 18 or 19 different bodies or people who have access to IRS statistics now?

Mr. BOWMAN. Quite a few. I never counted them.

Mr. GALLAGHER. I think it is 18 or 19.

Mr. BOWMAN. The way I have tried to express it is that the statistical Data Center would have the universal rule, if what I have intended were to come about, that it would not release to anybody information that could be identified with an individual, but many agencies would have data in the Data Center which were not subject to that same restriction. They might release the information, but the Data Center would not.

For example, now and again the Census Bureau may have a request from an agency for information about an individual. The individual may write to the agency and say, "It is all right. Release the information."

The practice there is still not to do it but to write back to the individual and, if he has lost a copy of his form, give him a copy of the form and let that information then be sent by the individual businessman himself.

Mr. CORNISH. This is handled by one set of Federal employees. Under the Center, actually we would have two sets of Federal employees involved.

Mr. BOWMAN. There would be two sets of Federal employees in the sense that the Center would have its own employees. Where the Center would be located, we do not know. It is also true that the other statistical agencies have people that are subject to the same requirements.

Mr. CORNISH. My only point was that you do get an increased problem in the control area.

Mr. KRUEGER. I think one of the points Mr. Bowman made earlier is that it is entirely conceivable that with the operation of a Center of this kind, the present practice whereby agencies now secure information from other agencies would be cut down, because they would not feel the necessity for doing that if there were a central capability of performing that kind of service.

Mr. GALLAGHER. Are we not really, getting right down to it, placing an unbearable burden of temptation on the keeper of the keys or the keeper of the safe—when we get down to putting all of this information into a Central Data Bank—that no one is ever going to want or have access to it?

Mr. BOWMAN. I think we have to impose close surveillance on what goes into the Center.

Mr. GALLAGHER. Assuming you do that, really are we not placing an unbearable burden of temptation on whoever is in control of that?



Mr. BOWMAN. I think not. I think if you expand it so it includes everything known; yes. If you do not expand it that way, I have the feeling, Mr. Gallagher, we might even be improving confidentiality.

Mr. GALLAGHER. Except for an elite who will be in possession of the information. It is that elite that the average citizen has a right to fear if you are going to put military statistics, crime statistics, census statistics, IRS statistics in there, and all that is retrievable if you just press the right button.

Mr. BOWMAN. That is why I said we probably have more carefully to scrutinize and specify what we are going to put in. I am not sure all of the things that people can think of as being there need be there for the Center to perform its most important function.

Mr. GALLAGHER. You have had a great deal of experience and I have great regard for your experience and for what you have done in the Government, but you know that once you start this thing as a "foot in the door," next year there will be someone else who will want to make use of your data bank. If it is that good, it will be expanded, and it will be expanded beyond a reasonable point.

Mr. BOWMAN. I think I agree, except on this point: You cannot stop moving forward in order to be sure that nothing bad will ever happen.

Mr. GALLAGHER. We do not want to impede progress. The computer is here to stay, and it can be a great source for good. We would hope that you are not underestimating the computer. I think you are. I think in centralizing this information in one giant computer, you have not realized the potentialities of the computer, because if you feel that you can control this kind of information and that the computer itself can make this decision, you are not being realistic. You are placing tremendous power in the hands of an elite.

I have a higher regard for the computer's capabilities than I think you have. I do not think we ought to abandon computers. We ought to utilize them as a source for good. But they also can be a source or mechanism for totalitarianism. I do not want to overstate the case, but if all this information is there and someone has total access to it, you can see there are a few dangers in it.

Mr. HORTON. Mr. Chairman, on that point I am not only concerned about the wrongdoer who may seek this information, but I am concerned about the attitude that will be created by this great source of information. The Internal Revenue Service information was very confidential at the start, but now it is available to States and, I think it was mentioned, 17 or 18 other agencies. You can justify every one of these uses of the information. There probably will be hundreds of uses of this central information that you can have that will be very good uses. Every one of us would agree right now that this would be a very fine thing. We want to stop criminals. We want to do this and that. This is a good reason for having this information.

But in the middle of that is a little guy who has now become a statistic. This can cause, I think, the loss of the privacy that he has, and this is what we are concerned about here.

Mr. BOWMAN. Mr. Horton, would you not agree on this point, however: If we make it clear that the center that I am talking about does not release information that can be identified with an individual, then it should be obvious, also, that information should not go into the center if the only reason for putting it there would be because people are interested in getting information about individuals.

Mr. HORTON. I wish we had the time to go into that. I think you have some very impractical impossibilities for curtailing the information that goes in, because you are not going to be able to get information that will be helpful if it is curtailed information at the inception. I agree with the idea. I think it is a good one. I also agree with your premise that we are not going to release anything except statistical information, but, as you have it, you can find out what happened to Frank Horton from the time he was born until right today just by pushing that button—everything.

Mr. GALLAGHER. It might never be released.

Mr. BOWMAN. I hope, however, what you say is not true for the data center that I am talking about. I did try to make this point. There are a lot of ways in which statistical data can be organized for collation purposes. If the data center is organized as I am thinking about it and does not release information about individuals, it would not organize the data that way. It would not be possible to press a button and get all the information about Mr. "X".

Mr. GALLAGHER. Mr. Bowman, we are going to try to adjourn here until 2 o'clock. We would like to finish up, if you do not mind. The kind of information that is anticipated to go into this disturbs me; credit survey, credit information, farm population data.

Mr. KRUEGER. If you are looking at this list as being the kind of information which would certainly go in, that is not true.

Mr. GALLAGHER. I have been looking at a list that has been pointed out as the possibility of going into this. Special census, metropolitan, Louisville, one-time survey; the census population of housing, 25-percent sample, population tallies.

There is all sorts of information that will be going in here of a very personal nature. Do you disagree with this?

Mr. BOWMAN. No. All those things have not been decided to go in. They would be available.

Mr. GALLAGHER. These are suggested items.

Mr. BOWMAN. No.

Mr. KRUEGER. This is an inventory of what is available in machine readable form.

Mr. GALLAGHER. This is on your 9,000 tapes now?

Mr. KRUEGER. It is more than 9,000.

Mr. GALLAGHER. Maybe it is more than 9,000. I am talking about the 9,000 tapes you are talking about putting into this as the possible material in the central data bank.

Mr. KRUEGER. No decision has been made of what goes into the center.

Mr. GALLAGHER. I would like to have you put that in the record. That would be helpful. This kind of information would seem to place people in peril and people would—I think, the average citizen who gives this kind of information knowing that it is instantaneously retrievable in a central data bank would walk with a certain amount of justified fear.

Mr. BOWMAN. It is available now.

Mr. GALLAGHER. It is decentralized. Once you centralize it you have a complete profile.

Mr. BOWMAN. Much of this is centralized.

Mr. GALLAGHER. If it is, then let us abandon what you are about to do if it is all that handy.

Mr. BOWMAN. Much of it is not.

The next point is even if it is centralized more than it is now, it is not available now and it would not be available when it is centralized so far as the individual is concerned.

Mr. GALLAGHER. You have told me you are not going to program out the individual's name before it goes in there.

Mr. BOWMAN. I have indicated that legislation to permit the disclosure of the information about the individual is certainly acceptable so far as the data center is concerned.

Mr. GALLAGHER. Right. The legislation is acceptable but the information is obtainable by those persons who have access to the central data bank keys.

Mr. BOWMAN. It is obtainable in the same way it is obtainable now.

Mr. GALLAGHER. Now he has to go to 25 areas and he has to say, "May I have this information?" and he can be turned down at every one of them and I am sure he probably is turned down. Again, I say that you are placing an unbearable temptation before whoever is in control of this central data bank not to use it for nonbenevolent purposes.

I would hope, Mr. Bowman, that you could give great consideration to this before you move forward on it.

Mr. BOWMAN. Certainly.

Mr. GALLAGHER. I appreciate your assurance that this will be submitted to the Congress for full consideration before even a pilot program is set up as a central data bank.

Do you have a question before we adjourn?

We are going to recess until 2 o'clock, Mr. Bowman.

Mr. BOWMAN. Would you like me to be back? I had planned to go off this afternoon but if you want me here I will remain.

Mr. GALLAGHER. Not as long as Mr. Ruggles and Mr. Dunn will be here. We want to thank you very much for being here this morning. I leave assured that you are full of good intentions but I leave with an uneasy feeling that they are incapable of accomplishment.

Mr. BOWMAN. More than good intentions, Mr. Chairman.

Mr. GALLAGHER. Thank you very much.

The committee stands adjourned until 2 o'clock.

(Whereupon, at 12:35 p.m., the committee adjourned, to reconvene at 2 p.m. of the same day.)

#### AFTERNOON SESSION

Mr. GALLAGHER. The committee will come to order.

The Chair would like to call Mr. Richard Ruggles and Mr. Edgar S. Dunn, Jr., to the stand.

Mr. Krueger, would you like to join in?

Mr. Ruggles, of the Yale University Department of Economics was chairman of the Committee on the Preservation and Use of Economic Data of the Social Science Research Council, which last year submitted a report urging the establishment of a data bank.

Mr. Dunn is a research analyst with Resources for the Future, Inc., of Washington, D.C. As a consultant to the Office of Statistical

Standards, Bureau of the Budget, he has written a review of the proposal for a national data center.

I might say that there appears to be a data wall, Mr. Ruggles, on the campus of Yale University. Professor Reich was here yesterday and took a dim view of your proposal. That is what we are here for—so that we can get all the views, to see that we ourselves can learn as much about the problem as possible.

I hope that this dialog in some way will create a climate of concern necessary before we take such a step.

Mr. Ruggles, would you proceed with your statement?

#### STATEMENT OF RICHARD RUGGLES, DEPARTMENT OF ECONOMICS, YALE UNIVERSITY

Mr. RUGGLES. The Congress is indeed wise to give serious consideration to the question of the individual's right to privacy, and the possible infringement thereof by the Federal Government, and even Congress itself. The danger lies not only in the massive files which are built up in the different Government agencies, but even in fragmentary information which may fall into the hands of someone who may use it to inflict damage or embarrassment to the individual concerned. The individual citizen is quite right to be wary of a government which can use information at its disposal to coerce, badger, or expose him in ways which are not based upon due process of the law. One of the most encouraging developments of recent years has been the increasing recognition that information obtained illegally does not constitute proper evidence, and that certain agencies of the Federal Government itself may have acted illegally in their attempts to procure such information. Thus wiretapping, improper seizure of records, et cetera, are now considered illegal in situations other than those directly concerned with national security. The use of confessions and the prosecution of offenders without adequate legal representation have also been called into serious question. By the same token, it is becoming increasingly evident that we must take steps to protect the information which the Government obtains from individuals and businesses in its normal operation. This, I gather, is the concern of this committee today.

The problem of disclosure of confidential information about individuals and businesses is not new. It has long been recognized that the information which individuals and businesses provide under law to the Bureau of the Census, for example, is confidential. This means that no other Federal agency is permitted to see or use the individual records, and even Congress itself cannot obtain census information on any individual or company. In fact, this confidentiality has been guarded so zealously that Congress and the other agencies of the Federal Government have been enjoined from obtaining from companies duplicate copies of those records which were submitted to the Census Bureau. The disclosure rules are meant to safeguard individuals so that they can feel sure that information which they give to the Census Bureau will never be used against them for such purposes as tax enforcement, antitrust, or congressional investigations. The disclosure rule has not been interpreted, of course, as preventing the use of census information for analyzing policy or providing in-

formation about specific groups, regions of the country, performance of industries, et cetera. In making tabulations of data, however, the Census Bureau carefully omits those classifications which might enable anybody to figure out information about individual firms or persons.

There are, of course, other Federal agencies which must by their very nature use information about individuals and firms for their operations. Thus for example the Internal Revenue Service not only must collect information about people's income and the taxes they pay, but this information can and should be used to prosecute tax fraud or tax evasion. Similarly, the Social Security Administration must process information about each individual over a period of years, recording his job status, family status, et cetera. This information is necessary for the determination of social security payments. Such use of individual information is of course justified, necessary, and legal. On the other hand, it is a real question whether tax returns or social security records should be turned over to other groups who may wish to use them for other purposes if the persons or firms to whom the records refer may individually be affected thereby. The question of the proper or improper use of information by different agencies is indeed a ticklish one, and procedures should be developed by both the executive branch and the legislative branch which will protect confidentiality and insure the privacy of the individual. In a great many instances, agencies may wish to obtain information not for operating purposes, but in order to make policy decisions and to guide future operations. Thus the Office of Education has a real interest in knowing how college enrollments may be expected to develop in the future. Those concerned with questions of poverty wish to know the dimensions and structure of this problem. In a great many of these instances, the agencies in question have contracted with the Census Bureau to provide them with such general information based upon sample surveys. In these instances, a disclosure and confidentiality rule must be developed which will protect the individual and yet yield the general information which is required.

In addition to the primary question of preserving the privacy of the individual, there are additional related questions which deserve careful consideration.

It is unfortunately true that because the United States possesses a highly decentralized statistical system individuals and, in particular, businesses may be required to fill out a large number of forms from different agencies asking for essentially the same kind of information. Business firms often complain that they spend a great deal of time and effort in making out reports to a variety of different government agencies, and that their life would be considerably simplified if the different groups could get together and make a single request for information which they would share. With respect to sample surveys, it is also true that a number of different Government agencies construct special samples to obtain information which they need for policy guidance, and because these samples are done independently little consideration is given to whether related work going on elsewhere in the Government might not be adapted to serve a number of different purposes at the same time. Careful consideration given to problems such as this might well cut down the bother, expense, and exasperation of

those from whom the information is obtained, on the one hand, and simultaneously reduce the cost and increase the efficiency on the part of the agencies collecting the information on the other.

Recently it has become increasingly evident that detailed individual information is much more useful for research purposes and thus for policy guidance than is aggregated and tabulated information. This is well illustrated by the tax model for estimating the impact of different tax changes, which was developed by Joseph Pechman at Brookings. Under arrangement with the Internal Revenue Service, a set of 100,000 tax returns was developed which represented a sample of all the individual tax returns in the country. The IRS took special precautions to eliminate identifying characteristics, so that the specific individuals could not be recognized. In evaluating the effect of a proposed tax change on various kinds of taxpayers and on total tax revenue, it was merely necessary to program the computer so that each of the 100,000 cases involved would be recomputed according to the proposed change in the tax law. By this device it became possible to evaluate the differences among various proposals, and to see how individuals in different situations might be affected. This research method has proved to be so successful that it is now part of the tax research program carried on by the IRS itself.

The same kind of research technique has recently been used at Yale by a student writing a Ph.D. thesis on the future economic status of the aged population. The primary material used for this research was the 1 in 1,000 sample of households prepared by the Census Bureau from the 1960 Demographic Census. Like the sample of tax returns, identifying characteristics which would permit the recognition of individual cases were removed, but data for each of the households in question was presented in considerable detail. Additional samples of data on household finances obtained from the Michigan Survey Research Center were used to construct a hypothetical income statement and balance sheet for each household, and data on such things as private pension coverage, labor turnover rates for various professions, et cetera, were obtained from the Bureau of Labor Statistics. On the basis of such information, a life process model was constructed, whereby each individual household in the 1960 sample was aged a year at a time for 20 years, taking into account mortality rates, possible job changes, projection of wage changes, and so forth. At the end of the 20-year simulation, the results were summarized to find out what the economic status of the aged population might be, in terms of the income distribution and the nature of those individuals who were at the poverty level. It is true that any single projection requires assumptions about the future social security payments, wage changes, pension coverage, et cetera. However, one of the major purposes of the simulation was to see the extent to which different kinds of assumptions mattered in the estimate of what might take place in the future. It is studies such as this that can help provide a basis for future legislation concerning many of our central problems. Although disclosure of individual information is not necessary, the use of detailed individual information is required.

What kind of satisfactory solution can there be to these problems? First, and foremost, it is essential to protect the individual from an invasion of his privacy and the misuse of information which may

damage or embarrass him. Second, it would be useful if we could reduce the tremendous flood of reports which are required from businesses and individuals at the Federal, State and local levels. Some simplification and rationalization here would considerably reduce the burden on the respondents, as well as the cost to the Government. Finally, however, in developing adequate disclosure rules we must be careful not to throw the baby out with the bath. Many kinds of analytical research require access to individual information, but this should not constitute disclosure in any meaningful sense. Techniques must be developed to preserve the usefulness of detailed information but at the same time insure the privacy of the individual.

Mr. GALLAGHER. Thank you, Mr. Ruggles.

Mr. Dunn, would you like to proceed with your statement?

Mr. DUNN. Yes, sir.

Mr. GALLAGHER. Please proceed.

#### STATEMENT OF EDGAR S. DUNN, JR., RESEARCH ANALYST, RESOURCES FOR THE FUTURE, INC.

Mr. DUNN. Mr. Chairman, I want to thank the committee for giving me the opportunity to discuss this issue with you today.

I think that the concern of this committee is a very legitimate and proper one and one that concerns me as well. I think it is very wholesome that we can discuss this issue together. I do think that there still remains some residual confusions that I should like to try to point up.

Much of what I have to say here is a recapitulation of my previous testimony and points made by my colleagues. However, I think there might be some merit in recasting them in a somewhat different way.

I think that it is important to recognize clearly that there are two basically different types of information systems: (1) there are statistical information systems, and (2) there are information systems that have as their purpose the generation of intelligence.

I might say that I use the term "intelligence" here with some misgivings and for want of a better term. My concern is that the term sometimes carries a certain amount of emotional freight, but I use it here only to make a technical distinction.

The distinction is basic. Intelligence systems generate data about individuals as individuals. They have as their purpose "finding out" about the individual. They are widespread and common and essential in our private and public business. They include such things as the medical records a doctor keeps to trace the changes in the well-being of his patient and the educational records the teacher keeps to trace the progress of the student. They include requirements essential to public administration, such as the licensing authorities' need to know whether a driver has legal vision, or the tax authorities' need for information to administer taxes.

Most of the intelligence information systems with which I have had any direct contact are restricted systems which have a specific administrative purpose and have not as their purpose the organization of intelligence about individuals into an integrated dossier of any kind.

It is conceivable that an intelligence system of this kind could be developed.

A statistical information system produces information that does not relate to the individual. It only identifies characteristics that relate to groups of individuals or populations. It has as its purpose answering such questions as these: What proportion of the residents of Appalachia possess income less than \$3,000? In what way does the mix of economic activities in New York City differ from Chicago? What activities seem to figure prominently in recent rapid growth of the Southeast, Florida, the gulf coast, and the Boston-Washington corridor? What proportion of the registered voters turned out in a recent primary and how were they divided between Republican and Democrat, urban and rural, white and nonwhite?

These are just illustrations but emphasize the range of the questions, which is infinite.

This is sufficient to emphasize that a statistical system is busy generating aggregates, averages, percentages and so forth that describe relationships. No information about the individual is generated.

No information about the individual needs to be available to anyone under any circumstances for the statistical information system to perform its function.

This distinction, it seems to me, divides the issue of personal privacy into two parts. The first part of the issue is reflected in this question: Can a statistical information system be developed and administered in a way that assures that it cannot be used as an intelligence system? I think the answer is an unqualified "Yes." That this can be done and done successfully there is no doubt. It has been done and done successfully for many years for those files that constitute the records of the Bureau of the Census. We don't have to speculate about possibility. We have demonstrated the possibility.

Let's see wherein this protection does exist and the ways in which it can be extended.

The protections are of two broad types: (a) those that stem from the design and technical characteristics of the system, and (b) those established by statutory and other legal restrictions prohibiting the release of individual data.

I think that it should be emphasized that such statutory restrictions do exist, as has been emphasized here previously. The statutory regulations governing the census files are a case in point. They have existed for a long time. They have worked well to prohibit illegal disclosure of individual data to anyone, and this includes the President of the United States. The details of the legal provisions in this area and their history are sufficiently complex that I am not qualified to review them, but Dr. Bowman here and the men on his staff and men on the staff of the Census Bureau can elaborate on this subject.

These legal prohibitions are supplemented by a formidable array of procedures embodied in the design of a statistical system and inherent in modern technology.

These legal provisions constitute the following:

The lay or public image of such a system is one of an automated monster with everybody's records that can be instantaneously retrieved by pressing buttons. There seems to be no awareness that the same technology that projects this frightening image has characteristics that can be and are utilized effectively to protect the sanctity of the individual record. Let me indicate some of the ways:



First, consider the technical form of the record. You see in my hand a piece of computer tape. This is the form of the record. You and I can't read it. Although there may be several advanced computer technicians in this room, they can't read it. To get any information out of this tape requires (1) a machine, (2) a codebook, (3) an appropriate set of instructions, (4) a technician. You see, when we put information into such a system we start with an inherent technical advantage over traditional records when it comes to protecting personal privacy. Information on paper in agency files can be very easily prostituted by any human being who can read and write and can be bought. The number of people who could get a specific piece of information out of a file of this type are very limited and very easy to identify. This facilitates the establishment of security procedures.

Second, I just said an appropriate set of instructions was necessary to get any information off the tape. This prompts me to make two points.

A statistical system will have the data formatted and stored and will have developed a set of instructions—known in the lingo nowadays as computer software—that facilitate the generation of statistical services. It may, in consequence, not be a very efficient system for intelligence purposes because neither the format of the record nor the standard instructions are developed for intelligence use.

However, there is a more important point. In addition to designing instructions that tell the machine what to do, you can give it instructions that can tell it what not to do. You can teach the machine to distinguish appropriate inquiries—statistical questions—from inappropriate inquiries—intelligence questions or individual data. Furthermore, you can go further than that. You can teach the machine to identify “trick” inquiries—either accidental or purposeful. That is, you can teach the machine to say, “This is a statistical inquiry but it is framed in such a way that the population or group you have defined contains only one individual or less than some specified number of individuals.”

Thus you can design a system in ways that prohibit any output other than a legitimate statistical analysis. The machine itself and the design system that operates it can be used to monitor the use procedures in ways that greatly increase the efficiency and assurance of security.

In addition, through system design and the kinds of instructions you give the machine you can teach it ways to disguise records. For example, you can replace all individual identification codes with a special security code that only the machine can use and which it uses only for the purpose of associating records. Under such circumstances, no one engaged in any part of the productive operation of the system would be able to identify any individual record without access to the set of translation codes that could be protected by special provisions of law and intensive security devices. The machine can also generate data useful for statistical purposes that are randomized or modified into “prototype” records that retain certain useful statistical properties while losing all identity as individuals.

We have had experience with all of these devices, legal and technical; they can be further perfected and extended as we attempt to improve the efficiency of the system for legitimate statistical use.

Indeed, one of the strongest arguments for an improved statistical servicing capability such as I have proposed, is the fact that some integration of procedure and regulation is important in making the best use of the technical safeguards that have and can be developed to assure personal privacy.

So much for statistical information systems. The second part of the issue of personal privacy is related to information systems directed to the uses of intelligence. The issue here is basically different. You can't ask, “How can an intelligence system be designed to protect the output of information about the individual?” That is what it is designed to do and it has many legitimate purposes recognized by law and society. We have already identified examples of these.

Whether or not as a matter of practice the rights of individuals to personal privacy have been adequately protected in these systems as they currently exist, I cannot say. My area of professional experience has never involved me with this type of information system. I would just like to make two observations.

(1) As a layman concerned with personal privacy, my impression is that actual and potential violations of personal privacy are more important and more serious in intelligence systems of this type. It is my impression that this is the area where investigation and reform are more drastically needed.

(2) It should be the concern of Congress that a general-purpose, public-servicing statistical information system should be protected from any use as an intelligence system. I think I have demonstrated that this has and can be done.

I have gathered the impression that there might be equal concern that intelligence systems likewise never be used as statistical systems. This is not the danger. There is not a symmetrical relationship. If you have a properly established and safeguarded statistical system, you can often gain by association—either permanently or temporarily—records generated for administrative purposes with the statistical system.

For example, the Internal Revenue records or the Social Security records produce as byproducts—that is, things that have to do with their public face, age, sex, race, and so forth—certain demographic characteristics of the population that can be used statistically for making good, economical statistical population estimates for the intercensal years when used in conjunction with benchmark data from the census. Again, it may be useful and desirable to make a statistical analysis of the characteristics of the population of public assistance recipients for evaluation of public policy, in order to evaluate some aspect of public policy.

Such uses of traditional administrative records can be accomplished without any violation of personal privacy. The essential point here is that in the association such administrative records would gain all of the protections of a properly organized statistical information system.

Thank you, gentlemen.

Mr. GALLAGHER. Thank you very much, Mr. Dunn.

You have both pointed out the need, though not very well, of the benevolent uses for which the central data bank is to be used. I was interested, Mr. Ruggles, in the programs that you pointed out as being successful and rendering a valuable service.



However, you, in your statement, said that the personal identification was removed. If it worked successfully there, why could not the personal identification be removed when you assemble data in the central data bank?

Mr. RUGGLES. For practical purposes, it is removed, but you must remember there are many ways of identifying people other than by their names. You take the wealthiest person in a small town. You don't need his name if you know he is the wealthiest person.

Various government agencies have been very wise in preventing the disclosure of individual information which might be used to identify the person even though the name is not given. As a matter of fact, in almost all the machine-readable records that are kept, the name of an individual never appears. An identification number is often assigned as has been suggested. The purpose of this is merely to keep the record straight so if you have an error, or you find something queer about a given case, you can locate it again in the record.

Also, suppose you are trying to match records. For example, you may find certain information is available in Social Security records, and certain information is available in Internal Revenue records and you wish to match it and add it to general demographic information in the census records. You need some way of linking these records. You are not interested in the person, but you are interested in bringing the information together.

Mr. GALLAGHER. In this we share no disagreement, on the necessity of bringing the information together. We are in some difficulty as to the necessity of bringing the individual's identification along with it.

If you say on the one hand it is not necessary, why does it become necessary to have it and centralize it?

Mr. RUGGLES. In many cases there is no other link to bring the information together except a person's name, or a company's name. I think the best case in point might be Internal Revenue records, matching the companies in the IRS data with the establishment data of the Census Bureau, it is necessary to know whether this is the same company. If there is no other linkage than the name, the name is a useful piece of identification. In fact, that is what we use names for generally, so I know who you are and you know who I am. It is a reference device, but it is only used as a reference device.

Mr. GALLAGHER. I am glad to know what we use names for, but I am trying to keep your name and my name out of a pool where if you wanted a promotion there wouldn't be a long trail as to why you shouldn't get that promotion if someone made a bad evaluation of you when you were going to school.

Mr. RUGGLES. Oh, I agree with you completely on that.

Mr. GALLAGHER. Then why do you feel names are necessary? Why should the individual then be identified in the central data bank?

Mr. RUGGLES. I guess it is the same reason why in a demographic census they collect people's names. If you took a census without anybody's names, you would be rather lost if you were studying migration and were trying find where a person moved to or something of that sort in the basic records. Were you to assign everybody a number, you wouldn't need a name.

Mr. GALLAGHER. We have done that already. I am talking about interchanging a person's name and a person's number.

What I am speaking of is, is it possible to establish such a central data bank without personal identification?

Mr. RUGGLES. I would think that for the vast majority of records this would be true, yes.

Mr. GALLAGHER. Would you recommend that be done?

Mr. RUGGLES. If each individual were assigned a number instead.

Mr. GALLAGHER. You are still leaving his fingerprints on your record.

Mr. RUGGLES. That is right. You must link records. In order to link records, you need some form of identification.

Mr. GALLAGHER. Then it is impossible to set up a central data bank without eliminating the identification of an individual?

Mr. RUGGLES. Or company or any other unit with which you are dealing.

Mr. GALLAGHER. Therefore, personal identification would have to be a significant factor in establishing your central data bank?

Mr. RUGGLES. I would say in order to process information, this is correct. There has to be some identification of a unit; otherwise you would have no idea whether you had duplicates in your system; you would have no idea as to whether you had coverage of the same group of people or the same establishments from different sources. It would be what was referred to here as an "electronic garbage disposal."

Mr. ROSENTHAL. There is no way to refine the input in this sophisticated machinery to eliminate that in advance?

Mr. RUGGLES. It would be very simple, as suggested here, to have a key that would assign a new identification number to each unit that you were using so that there would be no visible connection between any record you had and any meaningful identification.

Mr. GALLAGHER. But if you had a key, you could match it up?

Mr. RUGGLES. That is right.

Mr. ROSENTHAL. The key could control a group of individuals so we would have no way to break it down as to who the individual involved was.

It doesn't make sense to me at all. In other words, you say there is no way of eliminating the possibility of error being fed twice into the machine without individual information?

Mr. RUGGLES. You see, one of the problems in bringing records together that you are trying to match, is to make sure, that the records do in fact refer to the same unit. Supposing you have one set of records that tells about the employment of the husband. Another set of records tells about the employment of the wife. You have to bring the records of that household together so that you can see what effect certain factors may have upon the participation of the wife in the labor force. This means there has to be some method of identifying a household. This may be an address. In many cases it wouldn't be a name. The address is used in the case of establishments very often because many times a name is ambiguous. A name is only one piece of information about a unit.

Mr. GALLAGHER. The name is not ambiguous to the bearer of the name.

Mr. RUGGLES. Often it is spelled differently. Sometimes the first name is used instead of the last name or the middle name.

Mr. GALLAGHER. To me it is not an ambiguous situation for you to have all of my records. I know you have my records.

Mr. RUGGLES. That is right, but you would not know that your name was on them.

Mr. GALLAGHER. I would know there was a central data bank and 25 or 30 people would have a complete profile and they would have to go to a lot of trouble to assemble it, but they would not have to go to a lot of trouble for you to gather information you should not have.

Mr. RUGGLES. I believe though the things you would object to would not be that someone would know your name and address and things of this sort. What you would object to would be as you pointed out so aptly before, what someone said about you or what somebody reported, or a previous job or a previous offense, and I would presume that that sort of information would not be germane to these studies, and certainly not germane to a centralization of them.

Mr. GALLAGHER. Let's just stay on what Mr. Dunn has divided it into. We are not talking of the intelligence gathering center, but we are talking about a central data bank. A statistics gathering bank.

It is impossible to set this up efficiently or meaningfully without some personal identification of the people involved; is that correct? Whether by name or number or fingerprints?

Mr. RUGGLES. This would be true not only for this data center, but for Internal Revenue, for Social Security, for Archives, for Census, for all holders of records. Records have to be about specific reporting units; that is correct.

Mr. GALLAGHER. Therefore, these records that would be included in the central data bank would have a trail back to an individual; is that correct?

Mr. RUGGLES. Yes. I personally have never had the idea of a central data bank. That sounds too much like a data morgue and it may well be.

Mr. GALLAGHER. It sounds like a morgue to a lot of people.

Mr. RUGGLES. The Social Science Research Council report referred to a national data center which was a clearinghouse or cooperative enterprise with very strict disclosure rules, intended to be a service organization to produce statistical studies.

Mr. GALLAGHER. That definition could fit a bank, except for the last part of the sentence. While we call it a bank, some people might call it a data center.

Mr. RUGGLES. Yes, it is a data center. This is what was recommended by the Social Science Research Council.

Mr. GALLAGHER. Will you please tell me about the Social Science Research Council?

Mr. RUGGLES. It is a nonprofit research foundation that gives grants of money to scholars working in universities on various subjects. It has received money from the Rockefeller Foundation, the Ford Foundation, and it is essentially a grant-making organization—many of fellowships, scholarships, and so on; the grants are in universities. It is concerned with the orderly development of the social science disciplines themselves.

The group that was worried about this problem of information for the social scientist was worried about it in the same way that the scientists are worried about the development of their laboratories and laboratory equipment, and the people in the humanities are worried about the development of libraries.

The social scientists are, to an increasing extent since the computer has come in, concerned with obtaining the kind of information that will allow them to do economic analysis and other social research that will throw light on problems such as legislation, the development of our society in general.

Mr. GALLAGHER. Then you would be a user of the data center?

Mr. RUGGLES. Well, shall we say the profession would be a user?

Mr. GALLAGHER. Well, your organization.

Mr. RUGGLES. It is not my organization.

Mr. GALLAGHER. The organization to which you belong.

Mr. RUGGLES. The organization which requested that I chair a committee on this topic for them.

Mr. ROSENTHAL. How could they be a user of this?

Mr. RUGGLES. They would not be a user of it. They are essentially a professional organization.

Mr. ROSENTHAL. They are interested in fostering the interests of the profession.

Mr. RUGGLES. That is correct. Like any of the other learned societies.

Mr. ROSENTHAL. Did any Federal money go into this foundation which is then dispersed as grants?

Mr. RUGGLES. Not that I know of. They may have gotten some National Science money, though I do not know. I do not know about their finances.

In consulting with the various Government agencies, we found none of them were set up in a way to service either other Government agencies or people on the outside with an analysis of the kinds of information that they had within the agency. Thus, for example, Internal Revenue Service is set up to collect taxes and to deal operationally with the things related to the income tax. They are not organized so as to use as a byproduct the information that comes from the tax forms, for other purposes.

Mr. GALLAGHER. There are some who would disagree with you.

Mr. RUGGLES. Well, this is not their primary mission. They have set up a statistical group to publish the Statistics of Income, but only to a very limited extent do they do contract work from outside for either other Government agencies or people outside of the Government.

The Bureau of the Census has probably done the most work for other agencies and for people outside, but even here the number of requests they must turn down because their primary mission in collecting the demographic census and the other censuses is so large they are incapable of handling every request.

Mr. GALLAGHER. How many requests have been made by grantees of the organization to which you—

Mr. RUGGLES. I don't know. The Social Science Research Council?

Mr. GALLAGHER. Yes.

Mr. RUGGLES. The Social Science Research Council does not undertake studies of its own.

Mr. GALLAGHER. I said the grantees.

Mr. RUGGLES. I have no idea how many grantees there have been.

Mr. GALLAGHER. You have no idea of how many requests have been made for this information?

Mr. RUGGLES. No. I have no idea how many grantees there are. They give out a considerable number of fellowships a year. I just don't know.

Mr. GALLAGHER. Wouldn't that be a significant factor in asking for taxpayers' money? To create this center for the use of people to do studies? I would think it would be germane to determine how many people might want to use this.

Mr. RUGGLES. You can take three approaches to this. You could have a public opinion poll on it and ask people if they thought this was a good idea.

Mr. GALLAGHER. I would think that would be rejected.

Mr. RUGGLES. You could also go for some sort of referendum and essentially get the people's view on this.

Mr. GALLAGHER. I would like to run on that platform.

Mr. RUGGLES. I think I know how both these would come out. But this is not quite the point. If you are running a company and you decide that you need certain kinds of information in order to make intelligent decisions on costs and prices, you don't ask yourself, "After I get this information, how many people are going to use it?" You ask instead, "What is the usefulness of this sort of information to carry out the things that we want to carry out?"

Mr. GALLAGHER. You don't see a difference between running a company and running the Government? A person with a company can quit.

Mr. RUGGLES. Had the IRS in the case of Joe Pechman's tax model asked in advance, "How many people are going to buy this?" and he said, "Well, only one. I am the only person that I know of at the moment who is interested in it," then on the basis of demand it would obviously be absurd to cater to this whim. But IRS didn't use this as the basis for their decision. Instead, they asked themselves, "What is the merit of this proposal?" It is true that there was only one customer, but the research project was so successful and so useful for public policy and tax research the data were well worth making available. The IRS didn't do it merely because an outsider wanted it, and I am not suggesting that the data center be created to pander to outside interests, academic enterprises, or others. If the Government wants to run its own operation efficiently, it should at least study the merits of the matter and not ask merely how great is the demand in terms of the number of people wanting it.

Mr. GALLAGHER. I think the question is, How is it going to best serve the people, the citizen, and the taxpayer?

Mr. RUGGLES. But not individually.

Mr. GALLAGHER. Yes, individually, because I think you have to consider the cost-benefit ratio of what good this will do against what harm it will cause to the individual citizen.

Mr. RUGGLES. That is correct. With that I agree, but by "individually" I guess I wasn't speaking of the same thing. You should not ask how many people want this particular data, or how much will each individual asking for the data directly benefit from it, or even how many times will individuals use the data.

Mr. GALLAGHER. It could be potentially harmful to a great many individuals. I think it would be essential to establish just what would be the uses that this would serve.

Mr. RUGGLES. I would agree completely.

Mr. GALLAGHER. Has anybody given this any consideration, Mr. Dunn?

Mr. DUNN. Let me add a few comments on this matter of the uses. First, of all, I don't think there is really any way, as a matter of practical possibilities, to determine something that we might call the request incidence or frequency of requests upon statistical files in the Federal Government, and certainly no way of identifying specifically what they would be upon an improved service and capability of this kind.

Now, one of the principal reasons why this is so is because most of the distribution of the information which is generated by this process is undertaken through published documents like the Census Monograph, and so forth. These documents go out to all sorts of places, including standard library, reference, and referral services of various kinds. People come to these records and documents for all sorts of reasons without any way of ever tracing who tried to get into the record for what, you see, for what purpose. Here is the important thing.

Well, let me make another comment on the same point and that is this: To try to anticipate what the frequency of requests would be upon a system which would improve the service and capabilities of matching records in various ways and so forth is likewise impossible at the present time because it would be roughly identical with the kind of problems DuPont would face when trying to decide whether to produce nylon or not.

Mr. GALLAGHER. Not exactly. We already have data centers set up. I think you should determine what use is made of this information at this point.

Mr. DUNN. No, there is a distinction here and it is a very difficult distinction to convey and I am not sure I can be successful in conveying this point. May I take a few minutes to try to convey this and do it in terms of a simple kind of example?

Mr. GALLAGHER. Please do.

Mr. DUNN. I would like to start, first of all, from the point of view that is very fundamental here in the whole operation of using numbers for analysis in statistics, and so forth, and that is that no number will convey any information to anyone, without being associated with some other number, except as it can be related to other numbers and except as it can be identified with an attribute or characteristic of that number.

Let me illustrate: Suppose I were to write on the wall over here the symbol "2" and ask you what it means to you. You would probably say nothing. You might try to give it significance by saying that it is a number in an array of numbers that falls between 1 and 3; but if you would say that, you are already trying to associate it with some other number, you see. However, if you are really quick, you will immediately say "two what?"

Mr. HORTON. \$2 million for this system. That is what I thought the minute you said it.

Mr. DUNN. It might be that, but if I said "two apples," the information content jumps immediately. If I go on and associate the apples with the behavior of people and say two apples are eaten by the average person each week, the information takes another jump. If I

associate it with something else and say this compares with the per capita consumption of three oranges per week, the information takes another big jump. If I go further and say in spite of this national average, northerners eat more apples than oranges, then the information content jumps again.

Mr. GALLAGHER. Less acid.

Mr. DUNN. That is a ridiculous example in a way, but what I am trying to point out here is, a task of statistical analysis or any kind of analysis—and there is the backbone for policy administration—is a way of associating records or numbers that have attributes to private enterprise, governmental units and so forth. Unless you can start with this kind of information, you don't have any information. You don't have anything that tells you anything. It simply cannot be done without the basic building blocks which ultimately are responding units of some sort. The building blocks must exist because there is literally no way that you can generate a statistic without them.

If you think you can solve this problem by abolishing the building block and leaving some kind of residue of statistics, you are mistaken because in the process you destroy the basis for all statistics.

Now, let me qualify this—

Mr. HORTON. I think you have just indicated why we are very concerned about this.

Mr. DUNN. Precisely, and I am concerned about this too. I am just as concerned as you are. The basic reason I am concerned you see is, the kind of system we are talking about here that can be of service to the public decision process cannot operate—I emphasize “cannot operate”—unless it will protect personal privacy and security successfully.

Mr. ROSENTHAL. Mr. Dunn, following that point up, you have made much ado about the fact that we have operated successfully in the Census, and the information contained therein by the Bureau is rather inviolate and yet I have just checked the statute and it says:

The Secretary may, upon a written request, and in his discretion, furnish the Governors of States and territories, courts of record and individuals data for genealogical and other proper purposes from the population, agriculture, and housing schedules prepared under the authority of subchapter (2) of chapter V.

Were you aware of the fact that that information can be given out by the Secretary and that he has that discretion?

Mr. DUNN. I am not an expert on the legal provisions here. With your permission, Mr. Chairman, I would like to come back to the earlier point in a moment.

Mr. GALLAGHER. Could you respond to that parenthetically and then we can get back.

Mr. DUNN. I am sure there are hierarchies of legal provisions here and some overrule others.

Mr. ROSENTHAL. This is the only exception to the other limitations that the Secretary has placed upon him. These are exceptions.

Mr. DUNN. I know of no instance in which the census materials have been used—no other instance where it has been used as basically an information-retrieval device about information on an individual, except in the case of genealogical records and there, bear in mind, this is the individual requesting information about himself.

Mr. ROSENTHAL. That is not what the statute permits. It permits much more than that and the thing that distresses me is that it may

well be that the men we have in charge of the Bureau now are concerned about these things, but the fellow in the line after him may not be at all concerned about it and may disseminate the material and in a very unwise, abusive fashion.

Mr. DUNN. There is no point of difference between us at all on this point.

I made the point in my statement that I think we have had enough experience with the various kinds of technical and legal devices for protecting security, and they have worked sufficiently well that we know it can be done. We have demonstrated that it can be done.

I am not taking the position that all of the legal and technical provisions which have been in effect for all of the elements and fragments of the Federal statistical system have been adequate in their functioning; have been adequate for this purpose. I am saying that we have a base of experience which assures us that we can go on to review substantive process, technical process and legal constraint to assure that a statistical system can operate with proper safeguards and I would suggest nothing be done without that kind of assurance being generated in the process.

Mr. ROSENTHAL. That is the point I want to make.

You would agree with the testimony we had this morning that before Congress authorized the executive branch or appropriated funds for the executive branch to institute such an installation that Congress make sure there is adequate legislation to protect the constitutional rights of the individual.

Mr. DUNN. Precisely.

I am on record within the Bureau of the Budget, in the Director's office, with a statement to the effect, in writing, that before a formal proposal can be developed in terms of the substance of a system that would be proposed for implementation, that a very substantial amount of technical and legal staff work needs to be undertaken and to be accomplished, and that the first step is essentially something comparable to what the engineers in the Pentagon might call a “phase zero” study. We have to look at this first with the kind of resources which allow us to say “Here's what we want to do: Namely, provide for more efficient statistical services—and at the same time assure and strengthen—and I emphasize “strengthen”—“the protections against personal privacy, and here is what it is going to take to fulfill these obligations.”

Mr. ROSENTHAL. It seems to me that you fellows don't come in with clean hands if you hadn't considered that before you went to the trouble of printing all these documents and making this presentation. It would seem to me that that would be one of the first things I would have thought of before I went so far along as you people have. What you are doing now is reacting to the interest of a congressional committee. I have some doubts as to whether you would have done this had this inquiry not been held. It would seem to me your reputation would have been enhanced had you done this in the first instance on your own, of your own volition, and included it in all these documents, and books that you have prepared.

Mr. DUNN. I agree. I can only say in my defense in this regard two things: The first is—and this is really the only important point—that I was asked to come in and to review a series of procedural and



technical processes that are operating within the Federal Government as statistical services. They had nothing to do whatsoever with intelligence services of any kind, with dossiers, with any kind of system that would generate any information about an individual.

As a matter of fact, so far as I knew, this was a preliminary kind of a report for purposes of opening up the question for discussion, internal to the Bureau. I am not saying by that that it is inappropriate for documents to be released. I think it is perfectly healthy for the discussion to get out and in this kind of form. I welcome this and I applaud the work of the committee in this regard.

The only reason I didn't make a great issue over the personal privacy issue in the report was basically because I was operating within a frame of orientation which said we are talking about the way in which we improve a set of established statistical services, and we are essentially operating within a tradition where we just take for granted that we have got to protect the personal—

Mr. ROSENTHAL. I didn't mean to go into any narrow personal vein discussing these things, but apparently no one in the executive branch wanted to assume that responsibility and everyone wants to feel that they themselves are statistical technicians and their assignment was to stay within that particular role. As I see it, you can't have one without the other.

Mr. KRUEGER. If I may respond to this point, in examining this whole question, if we are subject to any dereliction at all, I guess it has stemmed from the basic tradition which we consider to be firmly established in the functioning of the entire Federal statistical system that the confidentiality of information reported to the Federal Government for statistical purposes must remain violate no matter what.

Mr. DUNN. Without it we have no system.

Mr. GALLAGHER. With that I agree, and I am happy to hear you say that, but one of the things we found in the confidentiality question business is that everybody seemed to take for granted that someone else will protect the confidentiality aspects and all of the people who draw a test of it feel that the judgment will be made at the Bureau of the Budget.

Now, I don't know how the Bureau of the Budget really got into the business of becoming the public conscience for truth and personality questions, yet they are the final authority on what is a good question to ask and what is a bad question, or whether or not a person's rights are being violated.

I don't know how the Bureau of the Budget got into this business.

Here we are once again taking for granted that the Bureau of the Budget will protect confidentiality. This is one of the problems that concerns us and I am very happy to hear all of you respond to what is a really serious problem to all of us.

Mr. DUNN. It is a problem to us.

Mr. GALLAGHER. But everybody takes for granted somebody else on the executive level will watch out as far as this problem is concerned. This is one of the problems and it is why we are in an era of corrosion of the individual's privacy. Ours is a very large Government and everybody else in the Government feels somebody else will watch out for this problem and this committee found there is not too great a concern for this, nor is there a center of interest where this specific problem is reviewed as to the interests of the individual citizen.

I don't think the Bureau of the Budget is properly geared for it. I understand they have 10,000 plus or minus questionnaires down there somebody is reviewing.

Mr. KRUEGER. It is a matter of record, as a matter of fact, in the hearings before the Supreme Court, in the case which Mr. Bowman referred to this morning, and it is in the decision of the Supreme Court where they took note of the concern of the Bureau of the Budget with this very problem.

Mr. GALLAGHER. I know you are concerned. The problem is I don't know whether this is where the jurisdiction rightfully belongs—to have everybody in the Government dump their problem into your lap and hope you will have the personnel to properly review it.

(Brief recess.)

Mr. GALLAGHER. The committee will come to order.

Mr. Rosenthal, being farther down the alphabet, will be here in a minute. If we could resume, it is your recommendations, in which we agree, that the safeguards that you have outlined here this afternoon be built into any program that will come forward as a result of the proposal?

Mr. DUNN. Very definitely. I would amend your statement only to say safeguards of this type, and to emphasize that we need in the process of designing such a program or system before its recommendation for implementation, to undertake staff studies which will pin down much more precisely the legal safeguards and modifications that might be desirable and required, and so forth.

Mr. GALLAGHER. What do you think of the proposal that I advanced on behalf of the committee this morning, of submitting your proposal to a further study with regard to the question of privacy and adequate safeguards?

Mr. DUNN. I think this would be a very productive effort, particularly insofar as the design of the legal protections is concerned and their appropriateness to achieve the objective in law that is desired. I would say, however, there are also some issues of information organization and production of substance here on which people of this type would not have a great deal to offer, and might easily misunderstand or confuse in some way, and would result in legal recommendations that were not effectively addressed to the control that was desired or legal provisions that inadvertently destroyed some essential technical characteristic of an operating statistical system. It seems to me both types have to work together.

Mr. GALLAGHER. We would get some Yale men.

It has been our suggestion that the refining process or the sifting process should be gone through before it becomes a hard proposal; that it go through a further committee which would not necessarily be opposed to the idea, but which would see that adequate legal requirements are built into the system, and that such a second committee of this sort be allowed to look at it from that standpoint.

Mr. DUNN. Are you asking me to recommend—

Mr. GALLAGHER. No. I said would you agree with such a recommendation and that it was necessary?

Mr. DUNN. I would not want in advance to state that one specific way of accomplishing this result or another is preferable, whether another review committee or through some kind of staff procedure or



what. I have not given any thought to this particular matter. I think it should be accomplished in some way, and this is something that should be considered very seriously.

In this connection, I want to make the point that while I functioned as a consultant to the Bureau of the Budget in generating this review, I am not currently a consultant to the Bureau of the Budget, and I appear here as a private citizen.

Mr. GALLAGHER. Yes. We appreciate your appearance.

Mr. DUNN. I present my views as a private citizen. I am not in a position to say that the Bureau of the Budget is or will or should follow these procedures in any official capacity as a consultant.

I would hope, however, the committee would not assume the fact I am no longer a consultant to the Bureau of the Budget is a reflection—

Mr. GALLAGHER. You have stirred up a lot of waves here already.

Mr. CORNISH. I think the chairman's suggestion this morning was when the Kaysen Committee had finished its work, that report be submitted to a panel of experts in constitutional law, and also in the fields of public policy that we have been discussing here this afternoon, to examine it strictly from that standpoint and to make some suggestions and comment on it so the Bureau of the Budget will have those views at hand when it makes its final decisions on whether to go ahead with this project and how it will do so.

Mr. DUNN. I am not sure I would want to say I have any firm opinion at the moment as to whether or not it would be appropriate to make the decision immediately following the submission of the Kaysen Committee report, or whether it might more appropriately and constructively come after the Bureau of the Budget had had time to do some additional staff work and to indulge in some internal thinking-out process following the Kaysen Committee report. I have the suspicion maybe the latter might be more productive.

Mr. CORNISH. In the course of the timing, certainly it should be before the center actually would be adopted as a firm proposal. I do not know exactly at what stage it might be, but certainly before that.

Mr. GALLAGHER. Mr. Dunn, we have now determined that it would be necessary to have some individual identification to a statistic; therefore, to a group of statistics. Each statistic must in some way be related back to a person, symbol, number, or name.

I would like to ask you one other question: You say this system is not an intelligence system. Would it be mechanically possible or impossible for this system to take on the characteristics of an intelligence system as the byproduct of a statistical system?

Mr. DUNN. I would like to answer that in the following way: First of all, I should like to give your concern in your question maximum support. As a matter of pure logic or philosophical premise, it is obvious that there is no system that can be designed by man that cannot also at least logically or conceptually be broken by man. To make any statement or take any position to the contrary is foolish. It cannot be demonstrated.

I am saying that I think we have accumulated in the area of statistical systems a very substantial experience with legal constraints and regulations and with procedures and practices that have worked, as a matter of fact, very well in protecting personal privacy in sta-

tistical systems; that having had this experience, we know these kinds of things can be made to work and we know these kinds of things can be extended in a number of ways and, as a matter of fact, strengthened through the application of the computer, that is, bringing the computer into the control process itself.

In the end, one simply has to say that we live in a system of law. Unless we have confidence that we can make a system of law work, we just do not have any recourse to anything.

Mr. GALLAGHER. Our system of law is conceived on a system of checks and balances. Are we not bringing about a vast imbalance here, and will we not have to rely on benevolent people using this for benevolent purposes?

Mr. DUNN. I do not think so.

Mr. GALLAGHER. Are we not creating the capability or are we not creating an instrument that can bring great imbalance? Suppose you had a nonbenevolent group in charge of a data center and then we threw Larry O'Brien's book on how to win elections into the computer, and then started to assimilate files on the 100,000 most likely delegates at the next convention. We would come out with some pretty interesting profiles and statistics, would we not?

Mr. DUNN. This is a complex question. There are several questions, as a matter of fact.

Mr. GALLAGHER. It is a complex issue. I am reducing it to a very simple political question.

Mr. DUNN. Let me answer in this way: First of all, I would agree with the statement Mr. Bowman made at the close of his testimony this morning. I think it quite appropriate and desirable that we not only think about controls upon the output, what can go out of the system. Here the basic notion is that no data about any individual can be released for any purpose—

Mr. GALLAGHER. Let us not talk about releases. Let us assume that it cannot be released.

Mr. DUNN. The other thing that is very desirable and necessary is to think about what kind of things it is legitimate to put into a statistical system. This is, I think, a matter of considerable interest and concern, and is another thing that needs to be thought very much about. Statistical systems as a matter of fact have never characteristically had very much information in them of a personal character, about the sex life of the individual or psychiatric interviews or things of that kind.

Mr. GALLAGHER. If you put in the MMPI—

Mr. DUNN. I am not sure I know what that is.

Mr. GALLAGHER. Minnesota Multiphasic Personality Inventory, which has a complete listing of your sex life, your religious beliefs, whether you love your wife or who you like to look at a movie with—all of this sort of thing. If you start programing that in the interest of statistics and press the button, you would have all the information you said would not be available.

Mr. DUNN. If it is not socially desirable to put this in, keep it out.

Mr. GALLAGHER. It is socially desirable. Otherwise, why did the Bureau of the Budget up until recently approve it as a test?

Mr. KRUEGER. We have never claimed specifically it would be socially desirable to put that particular kind of information into a data center.

Mr. GALLAGHER. Why would you use it in the first place?

Mr. KRUEGER. Someone had a very legitimate and important use for the information.

Mr. GALLAGHER. Then it is not impossible to think that someone would think it would be a very important use if we gathered up all of the people who took the MMPI and accumulated some statistics about how many people are happy with their wives or whether they believe in the second coming of Christ.

A lot of people do not believe in the first, so they are going to be discriminated against automatically by the computer.

Mr. KRUEGER. This is the importance of the question which needs to be given further consideration—what kinds of data should go into a center in the first place.

Mr. GALLAGHER. Yes, this is a very essential question, I should think.

Mr. DUNN. There is an additional point here, Mr. Chairman, which I think is worth some attention which has not come out in the discussion thus far, I believe. You have the data that are in there. Let us start with the assumption it is hedged about with all kinds of protection, procedurally and legal.

Mr. GALLAGHER. This is the 9,000 reels?

Mr. DUNN. No. The 9,000 reels, if I may, I would like to come back to, because there is a specific response on that which was not handled this morning, but that is not germane here.

You have data in the information system. In the first place, it is restricted data, restricted not only in terms of its availability but in terms of the kind of data you have in it. Let us even assume you have stuff in there that someone would like to get at. He has some motive for breaking the system, if I may put it in these crude terms. The simple truth of the matter is that it is just not going to pay him to try to break the system, quite apart from whether or not it could be done. Why? Because what has he got to do? He has to break the law. The penalties can be severe. He has to induce some key group of people, and maybe some of them are responsible to both Congress and the administration, or something of this kind, under the new procedures. They are going to break the law and violate their function.

After he goes through all of this, not only breaking the law himself but having a whole series of very impossible and improbable kinds of people to break the law, what is he confronted with? What is it that he wants? He wants some data about an individual. Why would he go to all this trouble and then, after he had gone through this kind of beltline or chain process, get to a file of several million records which must be searched through for this information?

Bear in mind, for statistical purposes this information is probably not organized with all of the data about the individual in one place. It is scattered all through the system, because that is the efficient way to organize, the format of data for statistical purposes, not the efficient way for intelligence purposes, but the efficient way for statistical purposes. He has to go through all this. If what he wants is a little bit of information about one individual or even a few individuals, it is much less risky and much more economical for him to go out and find out for himself, hire a private detective or drive by the guy's house—

Mr. GALLAGHER. He can do that now, but when you are centralizing, you are reducing the cost and increasing the risk.

Mr. DUNN. No. Presumably—

Mr. GALLAGHER. You just told me this could be converted into an intelligence system.

Mr. DUNN. Not if it is protected by law, and even then it would be a very costly thing. It would be a very improbable kind of thing to do.

Mr. GALLAGHER. Would it not be more costly to build in safeguards?

Mr. DUNN. I am talking about the safeguards. I am assuming the safeguards. I am not questioning the safeguards. I am saying we have to have those. I assume that you were raising the question, even assuming we have a protected system, don't you have a system where the stakes for getting into the system are not so high that safeguards might be brought into the system? I am responding and saying conceptually there is no system that cannot be broken, but if you have a system that is safeguarded appropriately, the cost of trying to break the system is astronomically beyond the value of getting an individual piece of data which could be much more economically and much less riskily gotten in some other way if what you want is some data about an individual.

You see, it just does not make sense to try to go into some great big file or record—for me, for example, if I wanted to find out something about Paul Krueger, to go into some big file or record where some limited number of characteristics, which might not be the ones I am interested in, are included with millions of other records, and try to prostitute a lot of people in the process, to find that out, when the chances are by hiring a private detective or going around and snooping a little bit myself or talking to a few of his friends I could find out practically all the things I want to know about him anyway.

Mr. HORTON. May I ask a question at this point, Mr. Chairman?

I think you have been here and heard Mr. Macy's article referred to in which he, the head of the Civil Service Commission, talks about the wonders of this new system and says that direct tape-to-tape feeding of data from one department to another may become common. Then he goes on to say how they have used this computerized file to get names to furnish to the President for candidates for presidential appointments. So they are already using this system.

Mr. DUNN. No, sir. May I make a correction? They are not using this system or any system like it. This is not a statistical system. It is an intelligence system. In my testimony I made a clear distinction and said I am not talking about that.

Mr. HORTON. You were talking about getting some information on Mr. Krueger.

Mr. DUNN. What I was saying is if I wanted to try to use a statistical system for intelligence purposes, it would be a very inefficient way to try to get intelligence.

Mr. HORTON. But you could do it. It is on the tape.

Mr. DUNN. You might be able to do it by breaking the law and getting six other people to break the law.

Mr. HORTON. We do not have any law on that now.

Mr. DUNN. I am saying you do not establish it until you do have.

Mr. HORTON. Could I ask a couple of questions, Mr. Chairman?

Mr. GALLAGHER. Mr. Horton.

Mr. HORTON. In this report of yours, Report No. 6, of December 1965, there is an appendix B which is referred to on page 6. You sug-

gest, under existing authorities, to begin work on these 9,000 tape nucleus archives identified in appendix B. I am not sure whose report appendix B is. Is this your report?

Mr. DUNN. It is my report; yes, sir.

Mr. HORTON. Appendix B is your report?

Mr. DUNN. No, sir; appendix B is not my report. That was produced by a gentleman from the Bureau of Labor Statistics by the name of Mr. Mendelsohn, who was on loan to the Office of Statistical Standards for the purpose of making this review at that time.

Mr. HORTON. He was working for you?

Mr. DUNN. That is correct.

Mr. HORTON. You adopted this appendix B?

Mr. DUNN. That is correct.

Mr. HORTON. As your report?

Mr. DUNN. That is correct.

Mr. HORTON. You went over it very carefully, I am sure.

Mr. DUNN. That is correct.

Mr. HORTON. In the report you talk about data bank requirements, and you talk about getting all these various reels that are now out in the other agencies, into this central statistical center.

Mr. DUNN. May I comment on that, because this is the 9,000-reel issue that was mentioned earlier, to clarify this point, because there is confusion residing on this point?

In my report I identify a number of functional areas in which some improvements in efficiency are useful and required. Many of these functional areas do not introduce anything new insofar as the issue of personal privacy is concerned. One of the areas that I was talking about in this area was the problem of archives. This was the fact that there exists at the present time within the Federal Government no established set of procedures or standards for determining what records in the agencies will be archived; that is, preserved in ways that can be used subsequently, nor any institutional means nor any standard funding process for seeing that they are protected.

One of the very serious problems we have had for some time is the fact that many very useful records are being destroyed inadvertently or unnecessarily.

Mr. HORTON. Wait a minute, Mr. Dunn. We do not have all night here on this—I wish we did have—to go over all this information. It says:

Data bank requirements. An estimated \$3 million to \$3.5 million and 3 to 5 years are needed to stock the proposed data center with Federal statistics now in existence.

You are not talking about any archives. You are talking about stocking the system. Then you say:

These data will probably comprise about 20,000 reels of magnetic tape. However, a Federal center could be stocked with a respectable volume and variety of data relatively quickly for about \$260,000 at the rate of about \$27 per reel. A bank of 9,000 tape files could be established in about a year. Such a course would provide a fairly representative selection of significant data including, for example, 750 reels of the Census housing data.

You are not talking about archives there, are you?

Mr. DUNN. Yes, sir. My language may not communicate it effectively, and I shall take the blame for that, but what basically was intended there, what we were talking about, was not the matter that we

would take some reels immediately and, without any previous authorization or any review of this process, set up a data center which would perform all these integrated functions. The thing which we had in mind there was that there is a tremendous amount of spoilage taking place in existing records because we do not have procedures for preserving them, and that very simply, within existing authorities and procedures, we could start with 9,000 reels or some specified number of reels—I do not remember the detail now—and put them under some provision that would assure they would be preserved and protected for future use, safeguarding them.

Mr. HORTON. That is not the substance of this appendix B. The substance here is to build up a library for the data bank. That is the basis of it.

Mr. DUNN. The other function—

Mr. HORTON. Now let me read you something else on page 3 of that appendix:

Agency representatives seems excessively concerned with the confidentiality question. Turning data over to a Federal center would be a breach of contract with respondents who have been assured that none but agency personnel would view their reports, it was said. I tried to convey the assurance that if a data center were established it would assume the obligation of protecting both the agency and the respondent. Since feelings on this matter run quite deep, some steps should be taken at the outset to vitiate them or discussions beyond this narrow consideration could founder. One constructive suggestion was made in regard to confidentiality. Mr. Robert Mencke, of the Securities and Exchange Commission, expressed the view that corporate concern dealt mainly with current affairs. It was his feeling that after a period of 5 to 10 years back, data could be exposed to public view without serious objection by respondents. There would be difficulty, perhaps, in applying such a rule retroactively, but a notice to this effect on future collections of data might serve to make the problem less troublesome in the years ahead.

How do you explain that?

Mr. DUNN. This is a report of a man who went out to interview the agencies and get their reactions, and he is reporting on their reactions. The reactions to me seemed to be perfectly sensible.

Mr. HORTON. You adopted this, you told me in the beginning. I asked you if you adopted that statement, and you said you did, and you put it in this report.

Mr. DUNN. Let me correct myself. I do not imply that any or all of the things suggested in the appendixes were included in the substance of my recommendations.

Mr. HORTON. On page 6, under subdivision C, you say under existing authority to begin work on the 9,000 tape nucleus archives identified in appendix B.

Mr. DUNN. That was to protect them from being destroyed inadvertently, not with the purpose of initiating a system.

Mr. HORTON. Now moving to the next consideration, this is going to be on that tape, is it not?

Mr. DUNN. What is going to be on the tape?

Mr. HORTON. All this information that comes from these various agencies and some in the future.

Mr. DUNN. Whether they are all in there or not, I would presume is subject to the review process which would be a part of the system design which says what is legitimate to go into the system and what is not.

Mr. HORTON. You very carefully pointed out the distinction between an intelligence and a statistical system, and this is very fine. It sounds fine. But the question that I have to ask has to do with the basic distinction between the statistical system and the intelligence system, because it seems to me that a statistical system is the same thing basically as an intelligence system. You can get intelligence information off a statistical system, can you not?

Mr. DUNN. There are records about individuals and individual attributes in a statistical system. They have to be there in order for it to work. It cannot be used as an intelligence system if it is properly safeguarded, and I maintain it can be properly safeguarded.

Mr. HORTON. When you say "properly safeguarded," I do not know any law that is going to apply to a machine. You detailed on page 4 how the machine could be set up to do this and not to do that. It could be set up so you could get personal information on an individual at a push of a button, could it not?

Mr. DUNN. You can design intelligence systems, yes, that will retrieve information and even dossiers, if you like, on individuals. This can be done.

Mr. HORTON. This system that you are talking about. I want to know technically, not legally but technically, can this system be set up so it could be converted into an intelligence system?

Mr. DUNN. Any kind of system can be set up to perform any purpose, yes. I would say we are not setting one up for that purpose.

Mr. HORTON. I understand you are not setting it up for that purpose, but what I am concerned about is the fact that you will have available in one place all this information that comes from the Department of Defense, the OEO, the Civil Service Commission, the FBI, from wherever and whatever source you want. All these magnetic tapes will be put in there. That information will be fed in one way or another.

Mr. DUNN. Not unless you decide all these things are appropriate to put in this in the first place.

Mr. HORTON. Have you decided what is to be put in the system?

Mr. DUNN. No, sir. That is not my role or function.

Mr. HORTON. Did you recommend anything with regard to that?

Mr. DUNN. I did not.

Mr. HORTON. All you recommended was setting up the monster, if you will.

Mr. DUNN. Whether it is a monster or not depends upon whether it is properly safeguarded.

Mr. HORTON. You said it could be called that by a layman, and I adopt that suggestion of yours.

Mr. DUNN. It would be appropriate to call it that if it were so designed to perform in that way.

Mr. HORTON. It could get all this information in it.

Mr. DUNN. It could be designed to do that.

Mr. HORTON. It could on the push of a button get this personal information about an individual.

Mr. DUNN. It could be designed to do that, but then it would be an intelligence system and not a statistical system. I would assume the object of the whole exercise is to safeguard against that in terms of both the legal constraints which can be applied and the substantive

and procedural processes that can be applied; using, as a matter of fact, the computer as an aid in the security process.

Mr. HORTON. I am sure I am not an expert on how Government works, but I know how these things generally happen. You start off with a report like this, and then the first thing you know, something else happens. The first thing you know, you are getting this personal information. This is a report, and I realize that you are not responsible for the legal ramifications of what happens once you make the report. I appreciate the fact that you are no longer a consultant with the Federal Government.

But you have made a recommendation for a central data center that could get all this type of information I am talking about, and you do not define in here, at least in your report, as I see it, the difference between intelligence and statistical.

Mr. DUNN. That is precisely the point I admitted to Mr. Rosenthal earlier in the testimony this afternoon. I regret that.

Mr. HORTON. Your recommendations are not along those lines. Your recommendations have to do with a lot of other things. That is, on page 4: "Manage the Archives records, develop referral and reference services, provide explicit facilitating services for users." I do not know who the "users" are, but it continues: "including file arrangements, cross-tabulation, extending output options, obtain translations, file modification, record matching." Can you identify which of these is statistical and which is intelligence?

Mr. DUNN. All are statistical.

Mr. HORTON. All of them?

Mr. DUNN. Yes, sir.

Mr. HORTON. Every one of them?

Mr. DUNN. Yes, sir. There was no time in the preparation of my report, consideration of any part of it, that there was any intention to address ourselves to the issue of intelligence systems at all.

Mr. GALLAGHER. Would it not be a problem to address yourself to that problem, so that it was not built into this system?

Mr. DUNN. I beg your pardon?

Mr. GALLAGHER. Would it not have been advisable to address yourself to that system in order to prevent it from having the conversion possibilities of becoming an intelligence system?

Mr. DUNN. This is something that may very definitely need to be addressed in the process; yes, sir.

Mr. HORTON. That is all.

Mr. GALLAGHER. Mr. Ruggles, how close has the Kaysen report been watched?

Mr. RUGGLES. I honestly do not know.

Mr. GALLAGHER. You are a member?

Mr. RUGGLES. Yes, sir. Professor Kaysen has taken a new job and in the process of transferring from Harvard to the Institute for Advanced Study, he has, I think, been somewhat busy and the committee has not met for a final time. The report is still in draft.

Mr. GALLAGHER. Tell him that he has overwhelmed the Congress up to now.

Mr. HORTON. Mr. Chairman, I did want to ask one question.

Mr. GALLAGHER. Yes.

Mr. HORTON. Mr. Ruggles, you are a member of this Research Council, Social Science Research?



Mr. RUGGLES. No, I am not.

Mr. HORTON. You work for them, is that it?

Mr. RUGGLES. No; the way the SSRC report came about, was that the American Economic Association requested the Social Science Research Council to undertake a study of the problem of the preservation and use of economic data.

I was asked by the Social Science Research Council if I would chair a committee to study the problem. This committee studied the problem for 3 to 4 years, conferring with the various government agencies and so on, and after that time brought out a report.

Mr. HORTON. That was the so-called Ruggles report?

Mr. RUGGLES. That is correct.

Mr. HORTON. The council is a private organization, is it not?

Mr. RUGGLES. Yes.

Mr. HORTON. Would the members be users of this type of information?

Mr. RUGGLES. No, they would not.

Mr. HORTON. They would not have any call on it?

Mr. RUGGLES. No interest or function.

Mr. HORTON. Do they call on the Federal Government now for any of these services?

Mr. RUGGLES. No, I would not be able to testify on this because I know nothing of their finances. I think they are mainly supported by Rockefeller and Ford and other foundations.

Mr. HORTON. I was thinking in terms of statistical information.

Mr. RUGGLES. No, they do not do any research work whatsoever.

Mr. GALLAGHER. I think you brought out before that grantees of that organization would make use of it.

Mr. RUGGLES. The SSRC gives scholarships and fellowships to scholars. It is like the other foundations and research groups that give away money.

Mr. HORTON. The grantees may be users of that information?

Mr. RUGGLES. That is right. In another connection, my wife and I are doing a Social Science Research Council Project at the Census Bureau on price-cost behavior of manufacturing establishments. This study is done under elaborate safeguards of disclosure, you will be happy to learn.

Mr. GALLAGHER. I am.

Mr. RUGGLES. In our research we never see the individual firm or establishment data. Nevertheless, the computer was used on establishment data for the first time in 1954. Methods were devised on the computer to match the individual establishment records to provide a continuous report for the same establishment for a period of time.

The purpose of this project was to study, for producers as a group, whether in a period of high demand, producers tend to raise prices more than their costs, or whether they tend to narrow their margins and are squeezed. This is important for problems of fiscal policy in the Nation, in worrying about how to dampen a boom, the rule of wage-price guidelines, and so on. There is currently a lot of interest in this sort of question. Although we are not interested in the individual establishment we had to get data for each establishment so that we could see how on an average they all behaved. This is the sort of work which the Social Science Research Council, together with the Census

Bureau, sponsored in a series of six or seven major monographs that attempted to utilize census materials for academic research. The money originally came from the Ford Foundation. This is a typical sort of thing they undertake.

Mr. HORTON. In your initial studies did you try to differentiate between this classification that Mr. Dunn has come up with now of today here, this intelligence system, and statistical system approach?

Mr. RUGGLES. No. We were very much aware of the problem of individual disclosure. This, as a matter of fact, was very, very evident to the committee right from the start and we did emphasize in the SSRC report that proper disclosure safeguards had to be provided.

Mr. HORTON. With this statistical system that has been described by Mr. Dunn, does it not have for all practical purposes the intelligence system built into it? All the information that is there is related to an individual; is it not?

Mr. RUGGLES. Yes, sir. I would say that what you point out is quite correct. There are really two ways of subverting this system. I worry about this and I think you do have to build safeguards. One way is, of course as you suggest, that instead of a benevolent group in charge you get somebody in who is not interested in the public good and he uses it for private purposes. We have to worry about this whenever we appoint a Secretary of Treasury, or whenever we appoint an Attorney General, or whenever any of these major appointments have been made.

Mr. HORTON. Excuse me for interrupting. I am worried about that but I am—

Mr. RUGGLES. You should be.

Mr. HORTON. I am also equally worried and more so worried about 5 or 10 years from now.

Mr. RUGGLES. That is right.

Mr. HORTON. The bureaucrat, as Mr. Packard described him yesterday, for the sake of efficiency and for recommendations to short-circuit the different problems that he might be up against says "it is important for us to have this information. There might be a good reason for it."

Mr. RUGGLES. This is the same thing.

Mr. HORTON. I am so happy we have one of these banks. I think it is wonderful. This is certainly the type of information we need. The point I am trying to make is this: I may very well in 5 years from now say this is fine and this information ought to be made available. What happens to the private individual citizen and his rights involved in that? That is the point I am trying to make. This does put this information into a central system and one of the points that has been made throughout our hearings has been that by having this information dispersed—birth certificates in one place, military records someplace else, and bank statements someplace else, it is somewhat safeguarded.

As a Member of Congress I am sure that the Government has some information on me about money that I have received or statements that I have made in connection with campaigns. All this is in different places at the present time. Now what we are proposing is to make it available in one central place where we can, regardless of how many safeguards we have, set the system up technically so that with the push of one button a card will come up and give all this information. That is accurate, is it not?



Maybe I have not covered all of it but that is possible, is it not?

Mr. RUGGLES. I do not think it is really possible to bring all those records together. If you want my frank opinion, I think that is just dreaming.

Mr. GALLAGHER. David Sarnoff said in his article that by 1976 a computer will be capable of making 400 trillion computations an hour, or 2 billion computations per hour for every man, woman, and child. (The article appears in app. 3, p. 298.)

Mr. RUGGLES. You do have dangers in the system and they are present. I think we have to concern ourselves with them. We have now Archives. Archives is a general central system. You have to realize this. They have under their direction and their control all of the IRS records. They have all of the presidential papers of Kennedy, Eisenhower, all of these people. They have all of the basic documents of the United States under their control. You are quite right. It is possible for these to be subverted. It may be that Archives should be broken up and we should decentralize it.

Mr. HORTON. But it might be possible, would it not, that it would be subverted, but for good purposes it could be made available. That is what I am concerned about also.

Mr. RUGGLES. That is right. I think that we have to worry about this and ask whether centralization of many of our functions is at all useful. I know the Congress worries about this a great deal. Perhaps we have too much Federal Government. I do not know.

Maybe even many of the powers that the Federal Government has should be broken up. For instance, maybe we should not have a single tax system.

Mr. HORTON. You are getting into something else.

Mr. RUGGLES. No, it is the same sort of thing.

Mr. HORTON. We are talking—

Mr. RUGGLES. Centralization brings power.

Mr. HORTON. We are talking about the preserving of the right of privacy of an individual and making it impossible for all of the information on him to be made available on just a snap of a finger. That is what we are trying to do. We want to safeguard him as best we can in this computerized jungle that he lives in at the present time.

Mr. RUGGLES. If what you are asking is, should we bring all the information from all agencies about every individual together, I would put an emphatic "No" to that. Then does it follow that we should make sure that no information about any individual ever gets together anywhere?

My answer to that would be "No." Obviously, the best position is somewhere in between. We do want to build these major economic statistics systems.

Mr. GALLAGHER. Mr. Horton's concern parallels the concern that you yourself in your statement point out:

One of the most encouraging developments of recent years has been the increasing recognition that information obtained illegally does not constitute proper evidence, and that certain agencies of the Federal Government itself may have acted illegally in their attempts to procure such information. Thus wire-tapping, improper seizure of records, etc., are now considered illegal in situations other than those directly concerned with national security.

Mr. RUGGLES. That is correct.

Mr. GALLAGHER. You recognize the fact that within the Federal Government, sometimes its enforcers do act illegally to obtain this information?

Mr. RUGGLES. That is right.

Mr. GALLAGHER. You have rightly stated that the courts have thrown this out. But on the other hand, we have all sorts of little cute devices now in the enforcing area where you can use illegal information and not admit where you get it, and not admit it as evidence, but under the broadening immunity statutes you can put somebody in jail forever, or to carry this to its logical conclusion, well beyond the point of the crime for which he might be committed on civil contempt of court if he has not admitted to the information being obtained illegally.

All this ties in with some of the concern a lot of people feel and we could very easily drift into a police state where you do have corruptions at every level. I am very happy to read of your awareness of these very problems.

These are the concerns of the subcommittee.

Gentlemen, on behalf of the committee I want to thank you very much for your appearance here and for your testimony and helping to air the problem a little more. You have made a real contribution to our understanding. With that, I want to thank you very much for being with us today.

Mr. DUNN. Thank you for giving us this privilege.

Mr. GALLAGHER. The committee stands adjourned until 10 a.m. tomorrow morning.

(Whereupon, at 4:30 p.m., the committee adjourned, to reconvene at 10 a.m. Thursday, July 28, 1966.)

## THE COMPUTER AND INVASION OF PRIVACY

THURSDAY, JULY 28, 1966

HOUSE OF REPRESENTATIVES,  
SPECIAL SUBCOMMITTEE ON INVASION OF PRIVACY  
OF THE COMMITTEE ON GOVERNMENT OPERATIONS,  
Washington, D.C.

The subcommittee met, pursuant to recess, at 10:19 a.m., in room 2247, Rayburn Office Building, Hon. Cornelius E. Gallagher (chairman of the subcommittee) presiding.

Present: Representatives Cornelius E. Gallagher, Benjamin S. Rosenthal, and Frank Horton.

Also present: Norman G. Cornish, chief of special inquiry; Miles Q. Romney, associate general counsel, Committee on Government Operations; and John Forsyth, special minority consultant.

Mr. GALLAGHER. The subcommittee will come to order.

The first witness this morning is Mr. Paul Baran, computer expert with the Rand Corp. In the opinion of the Chair, he is one of the few persons in the United States acknowledged as an expert in the relationship between the development of computers and the invasion of privacy.

Mr. Baran has received his B.S. degree in electrical engineering from the Drexel Institute of Technology; M.S. in engineering at the University of California; and his M.S. thesis dealt with computer simulations as an aided adaptive printed character recognition scheme.

While at Rand, Mr. Baran has also served on several ad hoc Department of Defense committees and as a consultant to the Institute for Defense Analyses and the Stanford Research Institute. The Department of Defense panels include those on the communications network switching and survivability of command and control.

Mr. Baran is author of many papers in the field of computers and communications and has lectured at the University of Michigan on computers in real time. He is a member of the Association for Computing Machinery and a member of the Institute of Electrical & Electronic Engineers, including its professional groups on computers and communications systems technology.

The Chair welcomes you here this morning, Mr. Baran. Would you proceed?

### STATEMENT OF PAUL BARAN, COMPUTER EXPERT WITH THE RAND CORP., SANTA MONICA, CALIF.

Mr. BARAN. Thank you, Mr. Chairman. I would first like to summarize my remarks and then delve more deeply into the reasons for my position. I do so in the role of the private citizen and not as a representative of the Rand Corp. or its sponsors.

I have been asked for my viewpoint on a centralized Federal data system, but I have broadened my response to direct attention to a larger issue: whether, in fact, we might not already be part way down the road building a system with all the obvious dangers of a single Federal data system, but without its clear visibility.

The problems of the invasion of privacy are, in my view, significant, and they will exist whether or not the central data bank is created by the Government. Individual data systems, both public and private, now being developed, can be tied together eventually into a network that will present essentially the same problems.

Every time someone proposes a system that handles embarrassing information, we are commonly reassured by the words, "Only those having a legitimate need to know will have access to the information." While the statement rings seductively of safety and is granted in good faith, its validity is sometimes overstated—particularly when such systems are interconnected. As one familiar with the inherent weaknesses of both our computer and communications systems, I am less sanguine about these assurances.

My remarks are intended to raise a healthy skepticism to this soothing sirup. While our computer and communications systems are fool-proof, they are not smart-proof. These systems are wide open to tampering by anyone sufficiently intelligent and motivated enough to take advantage of their weak spots.

My key suggestion is that we proceed slowly and cautiously to insure that proper safeguards are built into the systems from the outset. I do not propose delay as an obstructionist act, but only to allow the thought required to insure construction of the safety features required. But this takes time—and dollars. The full ramifications of the problem areas created are not yet well understood by the computer profession. And this is too important a decision to leave to any single computer manufacturer.

There is at present little financial incentive for any single computer system supplier to conceive and create new safeguards. Perhaps this may be attributed to the lack of sophistication of each individual user or perhaps no single user can demand extra new safeguards when almost comparable systems have been built in the past without such protection. I think it time that Government speak clearly as an advocate of the public interest in the future and initiate the improvements we desire. The first step would be to start considering some of the individual computer systems now being built from the viewpoint of sub-systems of a larger, overall system now under growth.

The men guiding the computer companies, the ones who must perform the detailed work of building safer systems, in my experience are among the most public spirited and enlightened in the Nation. But I believe that they will get on with the job more rapidly once their attention has been directed to the long-range implications of their babies.

#### PREFACE

It is a privilege to be invited to express my thoughts on the future possibility of the inadvertent invasion of privacy by the computer. I am highly pleased but, I must confess, surprised by the rapidity with which Congress is addressing itself toward the examination of this subject—one not yet generally recognized as being a problem by most in the computer profession.

This subcommittee is considering a specific proposal: to build a larger centralized data base for the Nation to reduce the costs of duplication of files and to provide more rapidly available information to those with legitimate need. The initial goals are valid and useful. But of common concern is the range of possible side effects. In general, I wonder about the potential threat to our historic right to privacy that could be endangered by a lack of appreciation for the present-day limitations of the computer and communications technology that could allow tampering of files by a sophisticated criminal, or a conspiracy, or even Government itself.

The initial questions are those of examining the proposed central file system, considering its weak spots, and creating a precise description of those safeguards required that are technologically and economically feasible. If the gap is too great, then clearly we should not build the system. But as a practical matter we should realize that eventual development is almost inevitable. We would do well to concentrate on the more constructive and larger issue of: How shall we control the development of the automation of all sensitive information files in order to best protect the rights of the individual and avoid a "1984" nation?

This may sound pessimistic, but if one can save money automating a data system it is only a matter of time until it happens. The only questions are: When? and How? I would like to add the further distressing thought that we may already be well along in the creation of the very system whose needs and dangers we are discussing today. This might sound bizarre, but consider the following line of reasoning:

Our first railroads in the 1830's were short routes connecting local population centers. No one sat down and laid out a master plan for a network of railroad rails. With time, an increasing number of such separate local systems were built. A network gradually grew as economic pressure caused the new links to be built to span the gaps between the individual routes.

We didn't start to build a nationwide telegraph network in the late 1840's; only independent telegraph links. But it was not long before we had an integrated nationwide network. Even the name, Western Union, recalls the pattern of independent links joined together to provide a more useful system.

We didn't start to build a nationwide telephone system in the early days of the telephone in the 1890's. Yet, today we have a highly integrated telephone network.

Such patterns of growth are not accidents. Communications and transportation are services that historically tend to form "natural monopolies." The reason is well understood. It's cheaper to share use of a large entity than to build your own facilities. Hence, if you were to look at the earth, say, from the far-off vantage point of the moon, it would appear that the growth of these integrated networks out of individual pieces is almost biological.

So much for history. What is of concern to us is that automated information files have the same properties as communications and transportation that causes the integrated networks to be self-agglomerating. It is cheaper to share the information by tying together independent systems than by building a very large number of highly duplicating systems without interconnection. But "information" can be too

treacherous a commodity to be widely disseminated with ineffectual controls. Even a little information improperly used can do irrevocable harm. Information is readily counterfeited. It can be quickly reproduced and widely transmitted very cheaply.

Today we can see the independent, private automated information systems being interconnected to form larger growing systems. The direction of growth is clear.

My thesis is this: Today we are already building the bits and pieces of separate automated information systems in both the private and government sectors that so closely follow the pattern to the present integrated communications structure that a de facto version of the system you are now pondering is already into the construction phase. It is in many ways more dangerous than the single data bank now being considered.

There is no culprit. No one set out to build our system. It's like little Topsy in Uncle Tom's Cabin who said, "Never was born. I spect I grow'd."

Mr. ROSENTHAL. Is it your point that these pieces are just being independently developed by various agencies and various groups, and, all of a sudden, they will merge by alleged necessity?

Mr. BARAN. Yes. We see this with independent credit systems built to cover small areas and then they find it is economic to cross-connect. We see this with airline systems, systems built for individual airlines are interconnected to swap information back and forth to get people reservations on other systems.

Mr. ROSENTHAL. So, the point you make is that even though the Government has not put their stamp of approval on building this system it is growing on its own because various groups are independently developing its starting points.

Mr. BARAN. Precisely.

#### THE CHANGING COMPUTER

Once when we spoke about a computer we meant a large room full of equipment. It was used by scientists to perform complex calculations and by businessmen to prepare payrolls, store inventory data, and similar generally repetitive tasks.

Both the form and the uses of computers are now undergoing radical changes. They have become so powerful, can store so much data, and process this data so quickly that it becomes possible to "time-share" a simple computer. You will be hearing more of this word "time-sharing" in the future. Time sharing means literally that. Many people have access to the single computer installation. The computer has so much capability and is so fast that it creates the illusion that each user has his own computer.

Users time-sharing a computer need not be in the same room or even the same building. They can be hundreds or even thousands of miles away. All it takes is a telephone connection to a remote electric typewriter-like device. Early examples of such systems include the local stock brokerage "systems" and airline reservation systems. In one system, over 3,000 separate stockbrokers around the country can push a few buttons to display the latest prices on any selected stock. Making and checking an airline reservation in a matter of seconds is

now so common that we forget about the good old days of manual reservation processing as we sat interminably in a drafty airport waiting to find out whether a seat on a desired flight was open. Parenthetically, these words were written before the strike and today's hot weather.

Mr. GALLAGHER. I don't think this will be needed any more.

Mr. BARAN. Right. This paragraph is inappropriate.

#### RECORDKEEPING

Simple recordkeeping, a mark of a highly developed economy, has been a prime area of development of these large computer file/communications systems. Much routine clerical work has been transferred to computers interrogated by humans. More people will have access to these systems at even greater distances from the computers.

#### SOME INDIVIDUALLY USEFUL SYSTEMS

Today we see time-shared file systems used to store insurance records. In a fraction of a minute it is possible for a clerk in another part of the country to check to see if one insurance renewal check has arrived.

To date such systems seem to pose no overt social problem. The information handled is not highly sensitive, and access is generally limited. But as new uses are being found for time-shared computers, a subtle change is beginning to take place.

In New York State, experiments are underway with a new criminal information record system. A police car uses radio to transmit automobile license tags into a central computer. If there is an outstanding warrant, the information comes back so quickly that a second police car located slightly ahead is able to stop the suspect. Preliminary tests have shown the system to be highly effective. We are all familiar with credit-check systems which use our driver's license numbers when cashing checks. This system is so fast that bad-check artists have been caught redhanded.

Seemingly such systems make a socially useful contribution. But let us look into the future and consider what their uncontrolled proliferation could mean.

#### THE TRAIL OF RECORDS IN A CIVILIZED LIFE

As we pass through life we leave a trail of records, widely dispersed and generally inaccessible—except with a great deal of effort and diligence.

Beginning with a birth certificate, we continue to accumulate hospital and medical records. We become deductions on our parents' income tax. In school we generate records of our grades, attendance, IQ tests, personality profiles, et cetera. Automated teaching will add to this recordkeeping. The volume of data recorded per child may be expected to increase even more markedly. After school we start accumulating employment, social security, and selective service records. We may get a driver's license. Most of us will apply for marriage licenses, and some will collect divorce decrees which will end in voluminous court records. If we are lucky, we will be

able to avoid arrest and jail records. We move from job to job in a mobile economy creating moving-company inventory records for our goods. Even as we move from place to place we leave behind short records of our airplane reservations and for some reason every hotel makes a ritual of acquiring and preserving the alleged names and addresses of its guests for posterity.

Book clubs and magazine subscriptions reflect our point of view and interests.

This is only a partial list. Play the game yourself and think of all the records you leave as you go through life.

#### WHY SO MANY RECORDS?

Behind all this creating of records is the implicit assumption that they will some day be of use. In order to be of use, there must be some means of interrogating the files to resurrect the information sought. Thus, we envision large families of systems, each individually useful. For example, an Internal Revenue Department investigator might wish to have immediate access to the tax returns of each of the associates of a man who is being audited to check for consistency of financial relationships.

A company may wish to have rapid access to its personnel files to know whether to give a good reference to a former employee.

A doctor may wish to trace the entire medical history of a patient to provide better input into a diagnostic computer.

The Veterans' Administration may wish to examine a man's complete military record and possible other previous medical records to see whether the ailment claimed as being service connected really is.

A lawyer for the defense of a man will wish to search for jail and arrest records, and possible credit records of all witnesses for the plaintiff.

Professional licensing boards may want to delve into any records to determine if an applicant has an unblemished character.

The military, in filling extremely sensitive positions, may even wish a record of all books borrowed by a prospective applicant to insure that his interests are wholesome and he possesses the proper political bias desired.

#### ACCESS TO THIS INFORMATION

Today it is difficult to gather such information about a prospective examinee. If one went through direct channels and asked most sources for their records about a person, he would most likely be rejected, if for no other reason than that the information is not readily available—cheaply. Even if the records were publicly available, the investigator would have to spend a great deal of time and effort delving through to discover pertinent data. Today, as a practical matter, if an individual wishes to obtain certain information about a person, he hires a private detective who charges a great deal of money and expends a great amount of time obtaining a little information from a portion of these potential records. The price for a fishing expedition for information is high and most of the fish are inaccessible.

#### THE IMPENDING PROBLEM

So much for the pleasant past. Consider the following argument: A multiplicity of large, remote-access computer systems, if interconnected, can pose the danger of loss of the individual's right to privacy—as we know it today.

The composite information data base may be so large and so easily accessible that it would permit unscrupulous individuals to use this information for unlawful means.

Modern organized crime should be expected to have the financial resources and access to the skills necessary to acquire and misuse the information in some of the systems now being considered.

We are concerned not only with the creation of simple "automated blackmail machines" using this information, but with the added implication of new powerful "inferential relational retrieval" techniques now being developed. Such techniques, when fully refined, could determine relationships of any person, organization, event, et cetera, to any other person, organization, or event.

Human beings, by their day-to-day need to make decisions using totally inadequate evidence, are prone to jump to conclusions when presented with very thin chains of inferred relationships. For example, merely plastering a man's name on billboards will markedly change the outcome of an election, if the other candidate's name is not equally displayed.

The use of private detectives to unearth defaming information on political candidates and their associates has become an increasingly prevalent feature of elections and is expected to increase in the future.

The cost per unit of dirt mined by unautomated human garbage collectors can be cut by orders of magnitude once they obtain access to a set of wide-access information systems now being developed. It is the sophisticated form of chain-relation blackmail that may be of the most social concern. The development of geographically widespread access systems uses communications lines to connect the users into the computer. There is a widespread belief—but perhaps not by this committee—that somehow the communications network used will possess a God-given privacy, but "it ain't necessarily so \* \* \*."

#### THE IMPACT OF COMMUNICATIONS UPON COMPUTERS

Using telephone lines modified to handle digital data, we are able to build an increasing number of geographically distributed time-shared computer systems. Many individual users are connected to a common computer base. Examples of such systems include airline reservation and credit checking systems for civilians and fancy display "command and control" systems for the military.

Simple recordkeeping, a mark of a highly developed economy, has been a prime area of development of these large computer file/communications systems. The development passes through several stages. First, much of the routine clerical work is transferred to a single large computer with few humans nearby allowed to interrogate the system. As time moves on, the number of people who are allowed to directly interrogate the system increases. Next, the geographical distance between the users and the machine increases. And eventually sepa-



rate systems are tied together to improve efficiency. The communication network forms one weak point in the system from the standpoint of eavesdropping and tampering.

Now, for some directions toward a solution.

Assume that not everyone is as honest and as trustworthy as ourselves—but is just as diabolically clever.

Appreciate that we will be dealing increasingly with complex and, hence, difficult-to-understand-all-the-details types of systems in the future.

The people who best understand the operation of each system will be computer designers who build the system in the first place.

The best time for applying fundamental safeguards is during initial system design. "Patchups" at a later date may be relatively less effective compared to good initial design that includes an awareness of the existence and importance of the problem.

We cannot expect and should not expect legislative action alone to substitute for good design. Even ignoring the lag of the legislative/judicial procedure, the detailed subject matter verges on or beyond the limits of legislative effectiveness.

As you know, laws and laws alone have been pathetically ineffectual in stopping the growth of widespread electronic eavesdropping and wiretapping. At most, the courts have succeeded in preventing police from using the same techniques available to the private detective or the criminal—or even a casual reader of an electronics technician's magazine.

While laws in themselves may not solve the problem, new legislation could be helpful in two ways: (a) laws outlawing certain practices will be of minor help in increasing the price of the act and making its commission less flagrant; (b) laws can be written so that potentially weak systems cannot be built unless adequate safeguards are incorporated throughout for the protection of the information stored.

This last direction is to me viscerally unsatisfying as it carries with it a built-in loss of freedom. The creation of another governmental agency peering over one's shoulder contains the possible dangers of bureaucratic delay and arbitrary conclusions based upon inadequate understanding of complex problems.

Historically, Government regulatory agencies start as highly effective bodies but lose momentum as the original personnel leave and their replacements come from the industry being regulated. Where else are you going to get competent people who know the business? The competence needed in a regulatory agency of this type is a too rare commodity.

If the computer industry is to avoid external regulation, then it behooves everyone who is involved with time-shared systems handling potentially sensitive information to start working, or at least thinking, about the problem of privacy. The computer industry should take the initiative and the responsibility of building in the needed safeguards itself before "Big Brother" is forced to do it himself and we may not be too happy with the way he might want to do it.

Safeguards, whether they be screens around moving machinery or circuit breakers, cost money. Every design engineer is reluctant to add anything that costs money and buys little visible protection. But it is time to start regarding such added costs as necessary costs—a price

to society for the privilege of building a potentially dangerous system. When you buy a new system of this type, plan to spend extra for safeguards.

This is not a new concept. We have, for example, been practicing this in the design of sewerage systems and in electrical distribution systems for some time. But, historically, it usually has taken an epidemic to build a local sewerage disposal system. It took a series of disastrous fires to get our electrical codes and possibly the recent Northeast blackout to start work on a better power grid.

The national geographical extent of the new data systems, their impact, and the investment are so large that the price of the "retrofit" after calamities occur may be higher than we need pay with preplanning.

To be more specific, what safeguards do I envision? Of course, we do not know all the answers yet. But, clearly, there are steps that we should be considering, including:

Provision for minimal cryptographic-type protection to all communications lines that carry potentially embarrassing data—not super-duper unbreakable cryptography, just some minimal, reversible, logical operations upon the data stream to make the eavesdropper's job so difficult that it isn't worth his time. The future holds the promise of such low-cost computer logic, so this may not be as expensive as it sounds.

Never store file data in the complete "clear." Perform some simple—but key controllable—operation on the data so that a simple access to storage will not dump stored data out into the clear.

Make random external audits of file operating programs a standard practice to insure that no programmer has intentionally or inadvertently slipped in a "secret door" to permit a remote point access information to which he is not entitled by sending in a "password."

When the day comes when individual file systems are interconnected on a widespread basis, let us have studied the problem sufficiently so that we can create sensible, precise ground rules on cross-system interrogation access.

Provide mechanisms to detect abnormal informational requests. That is, if a particular file is receiving an excessive number of inquiries or there is an unusual number of cross-file inquiries coming from one source, flag the request to a human operator.

Build in provisions to verify and record the source of requests for information interrogations.

Audit information requests and inform authorities of suspected misuse of the system.

This list is open-ended, and it is hoped that more suggestions will be forthcoming. But this will take much work.

Clearly here is an example of the trade-off between dollars and the type of society we want. We will face such decisions more and more often in the future.

What a wonderful opportunity awaits us to exercise a new form of social responsibility so that the advent of the new computer-communications technology need not be feared as we approach 1984. Rather, we have in our power a force which, if properly tamed, can aid, not hinder, raising our personal right of privacy.

If we default on exercising this opportunity and ignore the existence of such potential problems, the word "people" could become less a description of individual human beings living in an open society and more a mere collective noun.

It may seem a paradox, but an open society dictates a right to privacy among its members. Being aware of the potential problems is the first step in preserving this right. I sincerely thank the subcommittee for its efforts in this direction.

Mr. GALLAGHER. Thank you very much, Mr. Baran, for a very profound analysis of the problem as well as something that leaves us sitting here with our mouths open, as we consider in the light of the new technology we possess that "Big Brother" seems to be an infant in the year 1966 rather than 1984 if the technology we now possess were used for that purpose.

Mr. Baran, would you care to comment—I think you were here yesterday for part of the testimony of some of the proponents of the statistical data bank—would you care to comment on the feasibility of a system that could be designed for statistical data collection vis-à-vis an intelligence gathering system? Are they inseparable?

Mr. BARAN. The line dividing the two is an extremely fuzzy one. I think if one wanted to extract intelligence information from a statistical system he could. I could think of one example where a computer society conducted a survey—it has done this a few times—among its members, asking for information about salary, location, fringe benefits, length of service, and number of people employed. This information was sent to a very respectable accounting firm, magnetic tapes were prepared, and we were able to call up information of any type, saying on the average how much does a person make if he has been employed  $x$  years, with the following degrees? It soon became clear that we had to be very careful in protecting this tape because we could phrase our questions so that we could pick out people, even though there were no people's names on the tape. Yet the system described yesterday would have people's names and would have social security tags on them. This is a difficult problem.

On the other hand, the protection mechanism suggested yesterday that if the output of the inquiry covers such a small sample, that information is discernible as to precisely who is being examined, then the output is stopped, was also mentioned. This is the sort of safeguard I think one will have to see as a minimum. In other words, I believe the people yesterday were aware of some of the problems. It is unfortunate, I think, that they came in with a preliminary study and had not done all their homework on the safety part of the question. In specific answer to your question, you can extract intelligence information from a statistical system and get statistics from an intelligence system. It is just a little bit easier to use the machine for one purpose than the other.

Mr. HORTON. With regard to this so-called intelligence and statistical system difference, is it possible technically to design a system so that only statistical information could be utilized or be furnished and thus protect the so-called individual information?

Mr. BARAN. If you say I know all the questions I want to ask in the future, perhaps. But if you don't, that means you have to keep the information in raw form. This is the most efficient way of keeping it.

Mr. HORTON. I am assuming you are an expert in this field, the field of computers and what they can do. I am asking you from a technical standpoint whether or not it is possible—in other words, could we pass a law that would require—the construction of a computer that would only produce statistical information that would be foolproof insofar as individual information was concerned?

Mr. BARAN. "Foolproof" is a rough word. I think we could build safeguards to make it difficult. How effective they are, I think, requires a level of detail that we have not examined yet.

Mr. HORTON. The point I am trying to make is that I think any law Congress would enact to safeguard the right of individuals in this area would depend to a large measure upon the state of the art.

Mr. BARAN. That is right.

Mr. HORTON. With regard to the technical aspects, I do not think we have sufficient information to protect the private individual in the computerized systems.

Mr. BARAN. That is right. The technical art is changing very rapidly in computers. The speed of the computer is going up tremendously. The cost is coming down. The size of the memories is expanding very rapidly. As we look to the future we could probably see increases of size of computers—perhaps on the order of 10,000 times as powerful as today's computers.

Mr. HORTON. As a very simple case, if it were possible to pass a law that no computer system could have key A and that key A would be the key that would release personal information, there would be a safeguard. But short of that it seems to me we have a very difficult problem of enacting a law that is going to provide the type of safeguard that we are looking for.

Mr. BARAN. That is right. It is a very difficult problem to solve by law and law alone, because it is so difficult to implement the intent of the law.

Mr. HORTON. Have you given any thought to the technical aspect of how you could build in safeguards to protect private individuals' information?

Mr. BARAN. I think this is going to have to be done on a per system by per system basis. I do not think there is a general panacea. If a centralized statistical information bank is proposed, one would have to look at that particular system configuration very carefully in detail—in nuts and bolts detail—before making any statements.

Mr. HORTON. He could not pass a law for each system or each individual computer.

Mr. BARAN. That is the problem.

Mr. HORTON. How would we devise a law that would cover all computers? This is the problem.

Mr. BARAN. That is right. This is why it is too difficult a problem. All we could do in the way of law is to make misusing the information a crime. We would not expect this to be effective in itself—just increasing the price to those who would misuse the information. But there is no guarantee at all that this would solve the problem.

Mr. HORTON. I was not thinking so much of a crime as I was just to put in adequate safeguards that would prevent the misuse of information.

Mr. BARAN. It is a very difficult thing to do, because we are getting better and better information retrieval techniques. Even if you left the person's social security number off and left the name off, you could probably work backward and pose the question in a certain way and come up with the information.

Mr. HORTON. Were you impressed with the distinction that Dr. Dunn gave us yesterday regarding the intelligence system and the statistical system?

Mr. BARAN. I thought it was an excellent way to characterize their goals. In other words, what they wanted their system to do. Now comes the hard part—how do you build the system that will do what he says and not have the weaknesses that we are all aware of? I think the name of the game is really, how do we build safeguards into the system? We know it is just a matter of time until we get these systems. Clearly what was discussed yesterday was a very preliminary study. I think a lot more homework has to be done and a better definition of the precise safeguards reached. At that time I think we will be able to come up with more precise answers.

Mr. GALLAGHER. Mr. Baran, on the basis of that statement do you feel then that because it was so basically preliminary that it would be premature to attempt to build such a system at this point?

Mr. BARAN. I think the burden of proof ought to be put on the system proposal to show that no weaknesses such as you raised yesterday exist.

Mr. GALLAGHER. Has such a system been devised up to this moment?

Mr. BARAN. The military probably come closest to this. They have very sensitive information and have to transmit it from point to point. They go to a great deal of trouble and expense in keeping their information under control.

Mr. GALLAGHER. Isn't the control there basically one of cryptography?

Mr. BARAN. They have cryptography. They have certain clearances for the people who use the information. They have a system of need to know that is very rigid. And the penalties for abuse of the system are very high.

Mr. GALLAGHER. Is there a system now that can protect from a hardware security standpoint other than cryptography?

Mr. BARAN. I do not know how to answer that one. There are several weak spots in the overall system. You ask, "Can we do the job with other than cryptography?" Cryptography can be done on a software basis as well as hardware. In other words, the computer can generate its own cryptographic key. You would still have to watch your programmers and anyone having access to the information.

Mr. GALLAGHER. It basically is radiation transmissional interception.

Mr. BARAN. I think you have to assume that the other fellow will have access to the telephone lines. The information coming out of the computer has to be in such a form that he cannot decipher it. All we can do, I should say, is that we can make the price for breaking the information so high that it is not worth his while. It would never be foolproof. All we can do is raise the price to an eavesdropper to such a point to where he throws his hands up into the air and says, "The heck with it."

Mr. GALLAGHER. This automatically raises the cost of the system itself.

Mr. BARAN. Yes, sir.

Mr. GALLAGHER. Well, beyond the very simple figures that we have been talking about.

Mr. BARAN. I do not know how much.

Mr. GALLAGHER. I think part of the report was that they thought that they could start a system in the area of \$2 million. Is that even conceivable?

Mr. BARAN. I do not know. I have not seen their precise proposal.

Mr. GALLAGHER. Do you think the art is at such a state now that we ought to jump into this type of thing without having thoroughly thought it out?

Mr. BARAN. Heck, no.

Mr. GALLAGHER. We are in agreement with that statement for the very reason you have so ably pointed out. Are you suggesting incidentally to the line of questioning we are on now a possible need for the Federal Government to regulate the intercommunication of informational systems through the interstate commerce clause?

Mr. BARAN. I do not know. This depends on what the computer community does itself. In other words, here we are dealing with a new problem. It has descended upon us. We were not aware of it. All of a sudden we see these individual systems being built. If the computer community is able to understand what is happening and devise safeguards of its own, there may not be any need for the Federal Government coming in. This would be the most desirable course of action. If, on the other hand, we see systems being built without protection, and there is no interest by the computer manufacturers, then it may be necessary for the Government to step into this area.

Mr. GALLAGHER. On the question at hand of a statistical data center, what would you as a private citizen find desirable, or what would you expect the Government to do before we embark on such a course, before you would feel fully secure?

Mr. BARAN. I would just like the assurance that this information would not be misused. And by "misused" I mean passed over to others who do not have a legitimate need for the information.

Mr. GALLAGHER. Do you think there exists now a capability to give you this protection?

Mr. BARAN. I do not know. It is a very poorly studied problem. I speak to you with great ignorance today.

Mr. GALLAGHER. If you speak with ignorance think of the position we are in.

Mr. BARAN. That is right. We are all in the same boat.

Mr. GALLAGHER. Do you think that the people who are suggesting this proposal are beyond this point of ignorance that the rest of us share?

Mr. BARAN. I think we are all pulling on oars in the same boat.

Mr. GALLAGHER. Do you think that they have a broader understanding of computer capacity than you and your fellow experts have?

Mr. BARAN. I have not read their report, but from the remarks of yesterday they had not emphasized their examination of this problem. There is practically nothing to be found in the computer literature on the subject. Hence, I would think that they are not in any better position than the rest of us are in achieving a sense of happiness looking at the efficacy of the safeguards that we would like to see.

Mr. GALLAGHER. Would you agree with the suggestion of this committee that while the results the proponents are seeking are quite desirable, nevertheless the state of the art is at such an early stage that further study of the problem should be had, or there should be a continued study of the problem before we embark on a Federal statistical data system?

Mr. BARAN. I agree wholeheartedly with that statement.

Mr. GALLAGHER. In your statement here, I would like to read this part, where you were "highly pleased" and yet "surprised by the rapidity with which Congress is addressing itself toward the examination of this subject—one that is not yet generally recognized as being a problem by most in the computer profession." Do you think that the computer profession itself should start addressing itself in a more meaningful way to the problem of privacy and safeguards and protection?

Mr. BARAN. I do. I see a small but growing group in the computer profession beginning to become concerned about this range of problems.

Mr. GALLAGHER. Do you feel that before the Government would get committed on a course that we should allow the problem to develop and some ideas generate from the computer profession as to safeguards?

Mr. BARAN. Yes, I think we have the time. The problem we are addressing is a future problem. It is not something where we need an immediate answer today. We could hold off a little. We have the time to think and consider alternative approaches to this question of providing safeguards. Since the details are going to have to be devised by the computer profession, I think it is well that they be encouraged to start working along this direction. They work very well together. We have good industry standards on things like data communication, formats, and coding arrangements, and I am sure that something can be developed along this line in the matter of safeguards.

Mr. GALLAGHER. In view of that statement, that there is time to develop our thinking on this, conversely do you feel that this is not the time to get into a statistical data bank?

Mr. BARAN. If the report concluded that the need for such a center was sufficient, it might be well not to impede what they are trying to do. I think there are things to be done in parallel. I think that study of the data bank could proceed, but I would like to see a fair amount of parallel effort going into the safety aspects of the data bank.

Mr. GALLAGHER. I agree with you that the study should proceed. May I get back to this? Do you feel that it would be premature to embark on the construction of a data bank at this stage?

Mr. BARAN. From what little I have heard about it, I would think it would be premature to start on actual construction, because many of the safeguards are hardware-type safeguards.

Mr. GALLAGHER. Do you know anything about an Air Force research study of how to use computers to call up derogatory information on individuals?

Mr. BARAN. Not as such. I am aware of experimental information retrieval techniques that could allow you to ask questions such as, "What is the relationship between person A and person B?" "Have they ever worked at the same place?" "Have they ever coauthored a paper?" and draw connections between people or organizations.

Mr. GALLAGHER. Could such a system as that be built in an ancillary way to what we are now discussing?

Mr. BARAN. Yes. Once you have the information you can process it anyway you want to.

Mr. GALLAGHER. You also said that the centralization of the information would result in the loss of an individual's privacy as we know privacy today. Could you explain how privacy could be destroyed by such a system?

Mr. BARAN. Suppose we build our future systems without any safeguards at all and all information—this whole list of records that we accumulate during our life—is available in various systems. Let us suppose these systems are highly interconnected for the sake of economy. Then each and every one of these files can be separately interrogated, asking, "Do you have any information on so and so?" Do you have any information on so and so, and work on back and accumulate all the information you wish. Whether the information is centralized in one central data bank or whether it is spread around the country doesn't make a darn bit of difference. The result is the same.

Mr. GALLAGHER. Basically we could sum up, Mr. Baran, that we certainly should not attempt to impede the growth of technology but at the same time we should start devoting more time to building in safeguards so that this technology can serve man rather than subordinate him to its decisions.

Mr. BARAN. That is right. I do not think we are going to be able to stop technology. I think that decision is not ours. But what we can do is to provide all the safeguards we possibly can. This, I think, is the direction we should be going.

Mr. GALLAGHER. You feel at this moment that we have not utilized our resources in the direction of proper safeguards with regard to the privacy issue.

Mr. BARAN. That is right. As you know, it is a new problem. It is one that we have generally been unaware of in the computer field. It is so new we have just not built a large body of people concerned about this problem. I think the fact that you are holding these hearings will do very much to stimulate interest in seeing that such safeguards are forthcoming. The fact that you are holding these hearings will be sufficient to cause many of my colleagues to go back to the drawing boards tomorrow and start dreaming up ways—better ways—of protecting information.

Mr. GALLAGHER. We want to thank you very much for your contribution here this morning. Mr. Rosenthal.

Mr. ROSENTHAL. Thank you, Mr. Chairman.

In the broadest context, and one I think that perhaps I might understand and some of the American people might understand, it is really quite hard for us nontechnicians to visualize all of the manifestations of the proposal. But putting it in simple context, I seem to fall back on the proposition that it cannot happen here. I wonder if it ever could happen here, if that day came to pass. I think you cited some examples. Perhaps some incident that happened in high school might be fed into a computer. The type of car a person bought, whether he bought a motor scooter which apparently is popular with some people today. What book club he belonged to. What magazines he subscribed to.

What newspapers he subscribed to. Do you visualize that this type of information might some day be used by a less democratic government?

Mr. BARAN. This is the danger. This is the long-range danger. This is one way we could end up with a police state if we are not careful.

Mr. ROSENTHAL. I am not only worried about this being used as the mechanism to get to the police state. I worried about what would happen if the police state or anything resembling it came about because of any political changes in attitude by our people. Then this machinery would be available for us by administrators of that government.

Mr. BARAN. That is right. It would make for an extremely efficient police state if you had such a system.

Mr. ROSENTHAL. Is it the thrust of your testimony today that Congress could not today enact laws to prevent this from happening simply because neither we nor the scientists understand the technology sufficiently to protect society from these machines.

Mr. BARAN. Yes, but with some reservations. I think it is not quite that bad. I think we have some safeguards.

I think there are things that we can do. It is not a binary decision, that unless we do this immediately we are automatically going to end up with a police state. It is a shades-of-gray problem. In general, I tend to agree with your remark.

Mr. ROSENTHAL. I am not really making a remark. I am really inquiring because I find this different for a layman to understand. Is it your position that the technicians should be addressing themselves to the safety-valve features of the machine as much as the scientific development of the machine itself?

Mr. BARAN. Yes.

Mr. ROSENTHAL. And the scientific community has presumably not yet done that.

Mr. BARAN. It is just beginning. That is right.

Mr. ROSENTHAL. Is it because the scientific community is less obsessed than perhaps the Congress is with constitutional rights?

Mr. BARAN. I think there may be a matter of economics. If you could sell a computer system without safeguards and you add safeguards your price is going to be higher than your competitors. There is no economic reason why a computer manufacturer should develop these safeguards.

Mr. ROSENTHAL. In other words, you think we might be the victims of the economics of the situation.

Mr. BARAN. We could be. I think one of the outputs of this committee is making this range of problems more widely known, so that the people who buy computer systems become more sophisticated and ask for the safeguards. So when the bids come in from different computer manufacturers, the increased price will show up in all the bids and there would no longer be a differential factor.

Mr. ROSENTHAL. When you said the committee had an output, you are now taking us into the scientific computer.

Mr. BARAN. Yes.

Mr. ROSENTHAL. On page 17 of your statement, item No. 4, you made the statement, "We cannot expect and should not expect legislative action alone to substitute for good design." That merely supports what you just said informally.

Mr. BARAN. Yes.

Mr. ROSENTHAL. That the scientific community must address themselves by way of design to considering the problems that the subcommittee is considering, to wit, the invasion of privacy.

Mr. BARAN. That is right. Since this is occurring in both the private and public sector and is so widespread, I doubt whether you could sit down and write some laws and pass them that would solve the problem.

Mr. ROSENTHAL. Simply because the technology is not sufficiently advanced for those laws to be written today.

Mr. BARAN. That is right. That is one of the key elements.

Mr. ROSENTHAL. Thank you, Mr. Chairman.

Mr. GALLAGHER. Thank you, Mr. Rosenthal. You said we are in shades of gray. Would not a statistical data bank wash out the gray and replace it with black if we were to adopt it at this stage of the game?

We are talking now as a mechanism of totalitarianism or the police state that you mentioned.

Mr. BARAN. If we are to adopt it without protection possibly.

From the way the gentleman spoke yesterday, at least later in the afternoon they were aware of the existence of this problem. I think a suitable system emerging will have much more in the way of protection than the example that you have cited.

Mr. GALLAGHER. This would call for a great deal more study.

Mr. BARAN. A great deal more head scratching and hard detailed work on a specific system.

Mr. GALLAGHER. Thank you very much, Mr. Baran, for your appearance here today and your really outstanding contribution.

The next witness before the subcommittee is Burton E. Squires, Jr., visiting assistant professor of computer science, University of Illinois, Urbana, Ill. We welcome you here this morning.

#### STATEMENT OF BURTON E. SQUIRES, JR., VISITING ASSISTANT PROFESSOR OF COMPUTER SCIENCE, UNIVERSITY OF ILLINOIS, URBANA, ILL.

Mr. SQUIRES. Thank you, Mr. Chairman.

Mr. GALLAGHER. Will you please proceed? I saw you wince at a few questions.

Mr. ROSENTHAL. I did, too; I was not sure whether you winced because of the question or because of the answer.

Mr. SQUIRES. Mr. Chairman, I would like first to express my appreciation to you and to your committee in inviting me to these hearings. This is a very difficult and complex area, involving intimate knowledge in the fields of governmental and individual rights, psychometry, statistics, and modern technology. The task you have undertaken, to familiarize yourselves with the essentials of these fields, is indeed a formidable one. Through these hearings you have already made significant progress in protecting our right of privacy. You are now collecting data that will help all of us more reasonably to assess the full political implications of automatic data processing equipment. As you know, my specialty is in the field of computer science, and I will try to confine my remarks to this area. However, I feel quite deeply that we are here dealing with a technology that is as potentially dangerous and powerful as a nuclear explosive device.



It might be helpful to draw some analogy between the present technology in this area and the nuclear-device problem, but I fear the analogy may break down before it proves very useful. There is a very important difference in these problems. The nuclear device threatens to physically destroy our cities and perhaps our country in a rapid series of large easily recognized explosions. The invasion of the privacy of our citizens threatens to carry out a destructive mental process on a gradual, less perceptible scale, under the guise of causes that individually seem justified.

In preparing for this hearing I felt it appropriate to read again John Stuart Mill's essay, "On Liberty," in which he says:

The sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise or even right. \* \* \* Over himself, over his own body and mind, the individual is sovereign.

This committee has already heard extensive testimony regarding the collection of data concerning a person's thoughts, beliefs, emotions, and sensations. Professor Beaney and Mr. Speiser have testified—on pages 15 and 27, first session hearings—how such data can restrict the exploration of ideas, diminish intellectual curiosity, and restrain the free expression of thought. Mr. Mill's essay also develops these areas more fully. I would like, therefore, to extend these concepts as they relate to automatic data processing.

I cannot stress too much how important it is that this committee and our fellow citizens realize that a computer must be regarded as an information handling device rather than merely an arithmetic device. Perhaps the word "computer" should no longer be used, but rather some other expression such as automatic data processor, or information processor, or, as Dr. Peter Naur has recently proposed, datamation.

In addition to arithmetic operations, a computer can handle alphabetic and linguistic information. It can read, process, analyze, sort, store, and print such information at a phenomenal rate. A high speed computer memory now under development can read and write electronically at the rate of 16 million characters per second. A typical 300-page book contains about 1 million characters. The information storage capabilities are fantastic, although such a fast memory is extravagant in its use of space. It stores only 500 characters per cubic inch and it must be located within a very few feet of the central processor or "main frame" of the computer. Thin films and magnetic tapes are normally used for high capacity memories. A piece of magnetic tape about 0.0015-inch thick and 1-inch square, attached to a computer, can hold up to 3,200 alphabetic characters that can be read at rates exceeding 100,000 characters per second. This media packs information at a density of about 1½ million characters per cubic inch. Thus a building, containing 10,000 square feet of storage space 10 feet high, could conceivably store a book of information about every man, woman, and child in the United States. Specific information about any particular person could be transmitted along any given telephone line within a few minutes.

One often hears the remark that computers can do only what they are told to do. While this may be essentially true, it is practically false. Such a statement completely ignores the speed and complexity of problems that can be handled by a modern digital computer. It is, in a sense, like saying that an automobile won't take you anywhere you can't walk. Now, an automobile is about 15 times faster than walking. A modern computer is about a million times faster than paper and pencil.

However, a computer is a machine, not an animate being. As an automobile performs no better than the skill of its designers and driver, a computer performs no better than the skill of its engineers and programmers. Some of these programmers are extremely skillful and sophisticated. They can write programs which give the computer a kind of "artificial intelligence." In such programs the computer is allowed to operate in a simulated random manner, to evaluate the effects of these random operations, and to modify its own operating program. As a result, the computer can literally write its own programs for the direct solution of a difficult problem. By such means it is sometimes said that computers "learn." After a short time even the programmer has little knowledge of what the machine is actually doing, and he may be unable to predict the future behavior of the machine. The machine is able to learn because it was programmed to do so according to a specific learning theory. In this way the intellect of the programmer is still operating.

Computers now under construction will be able to process pictures as readily as present computers process linguistic information. A whole new era of information handling is upon us. It is quite reasonable to speculate that within the next 10 years computer terminals will be as commonplace as color television sets are today.

There can be little doubt that the establishment of a Federal data center could bring greater economy and efficiency to Government operations. It could do much more. It could make available to the executive branch immediate and up-to-date information summaries on all aspects of our national, business, and personal lives. Whether this can be done without violating the rights of individuals seems difficult at best and unlikely at least.

On the other hand, if the Internal Revenue Service is allowed access to the census data, and if the Federal Bureau of Investigation is allowed access to social security data, and so forth, or if these data are contained on magnetic tape so that they can be easily transmitted from one Government computer installation to another—as they are now or soon will be—then such a data center could come into existence in effect even if not in name.

Because of the rapid advances now being made in computer hardware and programming systems and in methods for handling large data files, the need for concerted activity devoted to the protection of the rights of the individual citizen has become increasingly urgent. The establishment of a Federal data center containing line item information about every man, woman, and child in this country is well within current technology. By line item information, I refer to information such as you might supply in filling out any form or questionnaire: an income tax, census data, an employment application, a security questionnaire, personality tests, etc. This information could be ar-

ranged or used in any manner seen fit by the persons having access to the data.

The motives and momentum of this trend are clear and must not be underestimated or misjudged. They are the good intentions of overly zealous public servants, and their goals are attractive: an honest income tax, instant credit, better employment opportunities due to better files on individual qualifications, more closely controlled production and distribution of consumer goods and thus presumably lower prices and fewer business fluctuations, and perhaps even better mating opportunities for our young people. For the first time we have a machine capable of taking into account and evaluating our individual characteristics, assuming, of course, that these characteristics are accurately known. We may not know how to use a computer for this purpose yet, but clearly it is possible.

We all recognize the necessity for the collection of certain kinds of information, such as credit ratings, for the operation of our complex industrial society. Although we may deplore this secret collection, most credit organizations today disseminate only the barest necessary information. Can we expect the same respect of privacy by a computer?

If the information is in the machine, and if an unscrupulous programmer knows how to get it out, it may be impossible to prevent his access to it. On the other hand, if it is so difficult for a programmer to retrieve the information that it is not worth the bother, the information is secure. How can we guarantee this?

We could, of course, try to avoid this situation in the first place. We could insist that laws be passed to make illegal the collection and disclosure of private records. Although such laws may serve some purposes, and they may significantly delay the effect, I do not believe that laws alone can prevent this information collection and retrieval—any more than laws can prevent automobile accidents. Laws merely specify how individuals ought to behave. They do not enforce that behavior. Our courts punish only individuals caught in the act of illegal behavior. The problem here is that our technological environment may make it difficult or impossible to apprehend certain types of criminals. The evidence of the information contained on a reel of magnetic tape, for example, can be destroyed in a fraction of a second simply by placing it in a strong magnetic field. This can be done as simply as pushing a button, just as today a telephone caller can prevent his call from ever being traced simply by hanging up.

If a felony results from the improper use of an automobile we arrest the driver, not the automobile. Likewise, if a felony results from the improper use of a computer, we must look to the engineer or programmer. Of course, in either case the fault could be due to mechanical or electrical failure of the device. The fault would be easier to detect if the device were designed in such a manner that its operation was unlikely to result in an accidental felony. Conversely, the successful operation of a computer depends upon those who build and maintain it and those who write its programs.

For many years our society has turned only to law for the solution of problems created by our technology. I think it time to expect technology to help solve some of the problems it has created. For some of these problems the basic scientific principles for their solution

already exist. We have already been successful in negotiating a nuclear test ban treaty with the assistance of a device for the detection of nuclear explosions. We must now ask our technology to help solve some of the other problems it has created.

I think the problem of the protection of individual rights in the use of a Federal data center is one of the problems which our engineers and scientists can help solve, at least temporarily.

I am not personally aware of any efforts in this area, but I think it is an area ripe for consideration. By this I mean it is an area in which it would be most profitable for some of our talented physical and social scientists to get together for a serious discussion of the issues and the potential solutions.

Over the years we have asked our talented young people to become scientists and to produce new devices for civilian and military use, and many of our people have contributed significantly in these efforts. We must now ask for the creation of devices that protect our freedom from the misuse of devices already produced. Automatic data processing equipment was not created without public cost and support; even less so will be devices to protect our freedom. Our forefathers did not gain freedom without risking their lives, fortune, and sacred honor. It appears unlikely that we can retain it without similar risk.

Mr. Chairman, that concludes my prepared statement. I would be happy to answer any questions that I can that you or the committee would like to pursue.

Mr. GALLAGHER. Would you care to comment on the proposal before this committee—that of a central data bank—whether or not we are ready and whether we should embark on such a proposal?

Mr. SQUIRES. Mr. Chairman, I have not seen this proposal, nor am I really aware of its economic and political implications in the sense that it is needed by the Government, so in view of my ignorance in this area, I don't think a comment would be appropriate.

Mr. GALLAGHER. Assuming that it could be helpful and that it is needed, from the standpoint of the technology and the lack of our young scientists—or our scientists, as you have pointed out here—to apply their talents to the problem of privacy and safeguards, would it be technologically advisable to embark on this course at this time?

Mr. SQUIRES. From the standpoint of the technology, as I see it, I do not believe we could embark on such a project without knowingly placing ourselves in very grave danger of violating the individual rights of our citizens. Whether this project could be undertaken in pieces in such a way that the amount of intrusion was minimal compared with the rewards that would be gained from such a system, I don't think I could directly comment.

Mr. GALLAGHER. I think you have summed up a problem that is in the area of our concern: that while we have made great advances in technology, there is an imbalance in regard to safeguards set up to protect the individual, his privacy, and his future sufficiently, in light of the advance of technology.

You state here:

However, I feel deeply that we are dealing with a technology that is as potentially dangerous and powerful as a nuclear, explosive device.

Would you care to elaborate on that sentence?

Mr. SQUIRES. One way this country could be turned into a police state is by means of a military force, conceivably. We go to great expense and effort to prevent this from happening. Our activities in southeast Asia are indicative of the cost which this community is willing to endure in an effort to prevent any such occurrence.

Mr. Baran has testified already this morning how the information contained in such a system could be used to create a very efficient police state. One might even say that, similar to the way coups occur in other countries, one would capture the building in which all this information is contained and one might be able to establish a police state just by virtue of having the information in his hand.

Mr. ROSENTHAL. Instead of the radio station, they capture the machine.

Mr. SQUIRES. You capture the information center because the information center is not only going to contain all the information you need to know to run the state, but very likely in the future it may contain all the information you need to control the flow of information.

Mr. GALLAGHER. And you agree substantially that a statistical data center could easily become a center of intelligence on individuals?

Mr. SQUIRES. Yes, it would be.

Mr. GALLAGHER. Do you see any way presently, in view of your experience, that the two could be separated?

Mr. SQUIRES. It would be most difficult. I would not say it would be impossible.

If the way in which you stored the information—this was alluded to by Mr. Baran—were not in raw data form, that is to say, if the way in which the information was stored was in the form of statistical summaries in which no summary dealt with fewer than say 100 or 200 people or some arbitrary number that would be sufficiently high to guarantee individual privacy, then of course at no time could you retrieve any such information out of the center because it wouldn't be in there. It was there only while it was being collected, and it was immediately destroyed.

On the other hand, this makes it difficult or almost impossible to ask at some later time a question which you had not previously anticipated asking at the time you made the summaries. Consequently, the information you collected is not nearly as useful as it might be because you have no ability to reorganize it in a different way simply because you destroyed it after you collected it.

Mr. GALLAGHER. Mr. Squires, on the question of liberty, where you quote John Mill's essay, you went on and your whole presentation gave the illusion that liberty could be fast disappearing in the new environment upon which we are now entering.

How can liberty be protected if these things will happen anyhow? Assuming somebody could get into the computer center?

Mr. SQUIRES. That is a very difficult and disturbing question, Mr. Chairman, and it is one which disturbs me and I frankly do not know the answer. I do feel that we ought to address ourselves to this question in a very real way in our society. I do feel it is a question to which we can address ourselves intelligently, but I don't necessarily feel that it is a question where we will be able to come up with answers without a great deal of effort.

Mr. GALLAGHER. What would be your suggestion to at least get on a track headed toward this end?

Mr. SQUIRES. My immediate suggestion would be to sponsor some sort of symposium in which prominent people and talented scientists and social scientists were invited to present papers on this subject and in which a reasonably small number of these people could get together and sit down and think up a lot of ideas.

Mr. GALLAGHER. You think this should be sponsored by the government, or some of the foundations or combinations of both? How can we get those people most qualified to think about this problem, to crank up and start doing something about it?

Mr. SQUIRES. Again the fact that you are holding these hearings I am sure is going to gather a good deal of attention among some of the foundations and perhaps just by virtue of what comes out of this hearing some of them might be willing to sponsor such a symposium. I do not know.

Mr. GALLAGHER. You, however, feel that it is a critical problem in the light of your testimony that we do start gearing our thinking in that direction?

Mr. SQUIRES. Yes, Mr. Chairman, I do. Personally I doubt that the computer community can be expected to come up with these safeguards on its own initiative without some help from foundations or from the Government or both, or some other sources.

Mr. GALLAGHER. Should this not be a joint effort with the social scientists? Some of the social scientists, by the way, are those who are strongly urging this statistical data center. I think there is a general unawareness in the community of the problem. Not an unawareness, perhaps just a lack of attention to the central problem that we are discussing here this morning.

Mr. SQUIRES. I am suggesting partly that it be a joint effort in that as a computer professional I'm really unable to evaluate the problems that the social scientist has. In a discussion between computer professionals and some social scientists, we might be able to come to some understanding as to what kinds of information the social scientists have to put in the machine, and what kinds of information they have to get out. Once we have this understood, the technology is in a much better state of being able to understand how to build some hardware that would have these protections in it.

Some of these protections can be built with software; that is, by programming efforts, but it seems to me basically it has to be built into the hardware, as Mr. Baran is also suggesting. I don't see how it can be built into the hardware unless the people building the hardware are aware of the kinds of safeguards that are really required. This is the reason I suggested a joint effort at this stage as appropriate.

Mr. HORTON. It seems to me also, Mr. Squires, at this point that it is well to differentiate between the control or the regulation in the public aspect—that is, through the Federal Government and governmental activity, and also in the field of the nongovernment.

As I understood the testimony, this is not a problem which could come about 5 or 10 years from now, but it is a problem that faces us right this very minute. We have before us just this one proposal with regard to a centralized data center, or a data bank. But there already is widespread use of this both in Government and outside of Government. So it would seem to me that these considerations of the protection of privacy, from my standpoint anyway, have a Federal interest, even though some of these computer systems are nongovernment.

Now, would you tend to agree with that aspect of it?

Mr. SQUIRES. Yes, I would.

Mr. HORTON. That we think in terms of protection of the individual's rights in the nongovernment aspect of our society in addition to the governmental aspect of our society.

Mr. SQUIRES. Yes, I definitely think so.

Mr. HORTON. I would like to ask you this: Are you in the technical aspect of computers? In other words, what is your specific field?

Mr. SQUIRES. Well, specifically I am working on programing systems.

Mr. HORTON. Are you an expert in the technology of computers, or are you more an expert in the field of the philosophy, or the use of computers, or both, or something else?

Mr. SQUIRES. I am afraid I don't really know how to answer that question.

Mr. HORTON. Do you design computers?

Mr. SQUIRES. No, sir, I do not design computers. I am presently at the University of Illinois. There are fellow scientists with whom I work who do design computers. My activities are more along the programing line, designing the programing systems, or software, as we referred to them in the testimony.

Mr. GALLAGHER. In other words, you engage in the computer sciences?

Mr. SQUIRES. That is right.

Mr. HORTON. You are more in the programing or use of computers as opposed to how they are built and designed, is that right?

Mr. SQUIRES. Yes, sir.

Mr. HORTON. Then I think this question is properly in your area: Do you have any suggestion as to how, programwise, this committee can make any recommendations or take any action which will protect the individual's rights in the use of the computers?

Mr. SQUIRES. I think some things can be done programwise, but they are somewhat minimal in that the information is stored in the machine somewhere.

One way of getting this information out of a computer is to do what is called a dump; what this amounts to is connecting the tape or whatever the storage device is to the computer and asking the computer to print out just exactly what is on this storage device. You then get a hard copy record of what was there. What you are asking is, is there a way to prevent a program from doing this? Strictly from a programing point of view, I do not believe there is.

Now, the information could be stored in the machine in cryptographic forms or it could be sorted and this should be done, as Mr. Baran has suggested, but even if it is in a cryptographic form, if a person knows the code or can invent a decipher for it and he has access to the computer, he can write a program to decode it. So I don't see programing alone as being the solution. I think it will have to be partly programing but a large part of it will have to be associated with the hardware.

Mr. HORTON. In other words, to define the areas with which we have to be concerned, we must consider how safeguards can be built into the technical system, that is, the mechanical system of the computers, and the whole problem of telecommunications and its ac-

cumulation and output of information. Then it also seems to me that we have to be concerned with building safeguards into the programs that are going to be involved in the use of the computers. Likewise, I feel we have to be concerned about the application of the system and the programs in the Federal Government or outside the Federal Government and build in safeguards in those areas.

Now, it seems to me that we haven't had very much information in any one of these areas until this subcommittee became interested in it, but I think you have indicated, Mr. Packard and others have indicated who have testified before us, that it is a very complex problem and that there isn't any safeguard at the present time in any one of these areas that will protect this interest that we are concerned about.

Mr. SQUIRES. There is some activity in the sense that people building time-sharing computers—particularly at MIT, for example—have developed systems whereby presumably a given user may not have access to another user's files. However, there are also other people in the field who claim that there hasn't been built a system that they can't break, so it is pretty primitive at the present time.

Mr. HORTON. In Mr. Macy's article in the Saturday Review of July 23, 1966, he makes reference to an integrated system and how this can be beneficial to the selection of personnel to fill jobs. He indicates some of this system has already been used in presenting candidates or prospective people to fill jobs at the Presidential appointment level. He also talked about direct tape-to-tape feeding of data from one department to another. It points out that this may become common.

I assume from what you have said and from what other witnesses have said that this is a possibility and in all probability is now in existence.

We also heard from the Government witnesses yesterday an indication that they would not furnish this intelligence information to various departments of Government, but I assume from what you and Mr. Baran said that it is possible to set up the system this way and possibly it is already set up this way.

Do you have any suggestion as to how, programwise, the Congress can deal with the restrictions on the use of this type of information from a central body of information to the various agencies?

Mr. SQUIRES. No, I don't, really. I read Mr. Macy's article on the plane, and read the quotation you just made and even made a copy because I was somewhat impressed by it. The idea of these existing channels is to me somewhat frightening. I can appreciate the problems that people running these systems have and I certainly appreciate their concern as a citizen that intelligence data will not be transmitted on these channels.

On the other hand, I do not see any visible means to prevent such transmission should the parties controlling the machines at any level decide to put such transmission into effect.

Mr. HORTON. It is a complex system that has already been built up to a large measure. I guess you would agree with that, would you not?

Mr. SQUIRES. Yes, sir.

Mr. HORTON. And much of this information is probably already available and it is now just a case of getting it in one central Federal system, or one central Federal place. Even though it is going to be proposed for statistical information only, the information will be there. Then the thing that bothers me is not so much that there may be the fellow who can steal the information or can tap in, eavesdrop, pick it up and disseminate it, but that it is going to be the call in 5 or 10 years from now on the basis of, "This is good for the country. This is something we need to know." In this way, these personal rights will be dissipated. This is a concern I have, too. In spite of the benefits of the system and the good that can follow from having this information, tapping it in and taking it for the public good will in essence do away with the individual's rights because all this information can be gathered in one place.

Mr. SQUIRES. I don't think the gathering of it physically in one place is at all a prerequisite for the existence of a Federal data center, or whatever you want to call it. Perhaps it is in terms of this specific proposal, but in more general terms, the physical location of this data in one place is not at all necessary.

Mr. HORTON. But this does create problems because, as I have understood it, the dispersal of this information at the present time is one of the safeguards that we now have. But if it is centered in one place, then it becomes a more serious problem, doesn't it?

Mr. SQUIRES. Yes, I would say it does.

Mr. HORTON. I think what we are concerned about now is the centralization of this, although I think our attention is going to be somewhat directed to the fact that some of this information is already gathered in some of these specific agencies and it is being used.

Mr. SQUIRES. It seems to me that one of the real problems we face as we gather more and more of this data is that it is difficult to make an intelligent and informed judgment on whether the need to know is as urgent as the need to protect the rights of the individual. When one queries general, popular sorts of opinion and points out the social advantages of knowing this information, the popular opinion will tend to say, "Let's have it," and, "What's all the secrecy about?" and, "Why does he want to protect this information. Is he some kind of a criminal, or what?"

It will tend to point the guilty finger at those trying to protect our rights of privacy.

Mr. HORTON. I thank you, Mr. Squires. I think you have made an excellent statement and you have outlined to us what the problem is in this whole area. I appreciate your coming before the subcommittee.

Mr. GALLAGHER. Mr. Squires, would you agree with Mr. Baran that any statistical data center, at the stage we are now in the computer business, that it would not be impossible to identify an individual if his records are programed into the data center?

Mr. SQUIRES. This depends not only on how the information is stored perhaps, but also on the kinds of access that are allowed to it. Let me suppose, for example, the information is stored in the system in such a way that there is no identity given. The social security numbers are completely divorced from the rest of an individual's file, for example—

Mr. GALLAGHER. You were not here yesterday. Mr. Dunn and Mr. Ruggles said that in the interests of statistical research it was necessary to have some identification of the individual—his name or his number. With that being part of his file, would there be any way in which he could be fully protected if someone wanted to get to the information as regards his personal characteristics?

Mr. SQUIRES. Fully protected, I would think not. Substantially protected, quite possibly.

Mr. GALLAGHER. Let's say the supreme commander of the control center wanted information. Could he get it if that information is in a statistical data center, as opposed to calling it an intelligence data center?

Mr. SQUIRES. If I understand your question correctly, what you call a statistical data center and what you call an intelligence data center does not differ in terms of the kind of information that is stored therein, is that correct?

Mr. GALLAGHER. Yes, that is the point I am trying to make.

Mr. SQUIRES. And if the information is stored in there and we do not have safeguards to prevent its being disseminated—and it appears to me at the present time we do not—therefore, my answer is clear. It could be gotten out.

Mr. GALLAGHER. Another intriguing question I always find of interest: Since the IRS has now set up a central data collection service and now that we have the potential of erasing from the computer's memory and truly making a person an "unperson," would it be possible for a skilled computer expert to make himself a nontaxpayer, by programing himself out of existence?

Mr. SQUIRES. That is a very interesting question. I suspect that it would be.

Mr. GALLAGHER. Therefore, by sending in the wrong card or the right card, or the wrong answers, he could be eliminated from existence from the rolls of the IRS.

Mr. SQUIRES. That seems to me quite reasonable.

Mr. GALLAGHER. And we have a "Ministry of Truth" where a person could be programed out of existence and become an "unperson."

Mr. SQUIRES. He would no longer be on the tax rolls at least.

Mr. GALLAGHER. Similarly, a person who might conceivably be in a data center and who would not want to have the information there—he too could be snapped out of existence by a proper program if such a program were set up, so there are two sides of this coin: One, a maximum amount of information on an individual and two, a person with the right key—or, as I said yesterday, the janitor who might have a key over there so he could get in to sweep up—he too could become an "unperson" on the rolls of the Government.

Mr. SQUIRES. This is conceivable.

Mr. GALLAGHER. Mr. Squires, I wish we could go on, but time is running. We want to thank you very, very much for your excellent presentation here this morning.

Mr. ROSENTHAL. I am going to pass and refrain from asking any questions in view of the fact that the House is now in session.

I have visions of Charlie Chaplin in "Modern Times," coming in being the janitor, coming in late at night and eliminating himself.

Thank you very much.



Mr. GALLAGHER. We thought it would be useful to hear about an experience of a State setting up a computer system. Mr. Robert Gallati, as director of the New York State Identification and Intelligence System, has volunteered to come before this committee to give us the benefit of his experience with such a system. The Chair would therefore like to call Mr. Gallati.

**STATEMENT OF ROBERT R. J. GALLATI, DIRECTOR, NEW YORK STATE IDENTIFICATION AND INTELLIGENCE SYSTEM; ACCOMPANIED BY ELIOT H. LUMBARD, SPECIAL ASSISTANT COUNSEL FOR LAW ENFORCEMENT TO GOVERNOR ROCKEFELLER, AND EDWARD DeFRANCO, EXECUTIVE ASSISTANT TO THE DIRECTOR**

Mr. GALLATI. Thank you very much, Mr. Chairman. I would like to say that it is quite a privilege and an honor to be here and to testify before the Special Subcommittee on the Invasion of Privacy of the House Committee on Government Operations.

I believe I might tell you a little bit about the New York State Identification and Intelligence System from the perspective of one who came from and was a member of the New York City Police Department for a little over a quarter of a century.

Mr. GALLAGHER. My father was a member of the Bayonne Police Department.

Mr. GALLATI. I had the position of assistant chief inspector, chief of planning, on the New York City Police Department when I was first invited by Mr. Eliot Lumbard, who sits on my left as counsel today, to participate in an advisory committee meeting to consider the creation of an information-sharing system for the agencies of criminal justice in New York State.

Among those on the Advisory Committee, besides myself and Mr. Lumbard as chairman, are representatives of the New York State Association of Chiefs of Police, the New York State Sheriffs' Association, the New York State Police, the head of the State parole board, representatives from probation, representatives from the New York State District Attorney's Association, and members of the staff of the department of correction, and members of the staff of the judicial conference. Represented are the six major branches or functions, if you will, of the administration of criminal justice. The police, prosecutors, criminal courts, probation, institutional services, and parole. We are joined by members of the Systems Development Corp., a "not for profit" organization engaged in various system development endeavors and at that time—and I believe subsequent thereto—engaged largely in Air Force contracts relative to the SAGE system. We discuss matters relating to the need for information sharing.

As I am sure you are aware, New York State is a large State and has tremendous geographic area, which causes a diversion in a sense, of the various jurisdictions on the horizontal, or geographical area, if you will.

Mr. GALLAGHER. Coming from the other side of the river, I sometimes question the jurisdictional rights, especially the claims to the "offshore islands" of New Jersey.

Mr. GALLATI. We welcome our sister State of New Jersey. As a matter of fact, we had last night the annual conference of the New

York State Chiefs of Police and we had as a guest Chief Harry Knowles who is president of the New Jersey State Chiefs of Police.

Mr. HORTON. You are covered well in the hearing today.

Mr. GALLATI. In the State of New York there are some 70 million files in the various agencies of criminal justice. We are aware of the fact too that in the operation of criminal justice in New York State, as well as elsewhere in this country, the police, the prosecutors, the criminal courts, probation, correction, and parole agencies, deal with the same person through a process, a criminal process, a process of Justice, and the same basic goals are shared by these various agencies for the social good.

In the process, there is a redundancy of investigation. There is also a redundancy of recordkeeping.

And so on the vertical level, if you will, we were made aware in our deliberations of the unitary nature of the process administering criminal justice and, therefore, the compelling conclusion that perhaps a unitary system of compiling and distributing and sharing information would be something to be seriously considered. Information is the raw material of criminal justice action.

The questions that were raised were, of course, what information should be brought together and shared, and we went into a considerable study of these matters and many things were accepted and many things were rejected as appropriate or inappropriate for storage in a central information-sharing system, be it computerized or manual.

It was thought that the tremendous amount of data which would be part of a system of information sharing such as contemplated by the advisory committee, would require the advances available to us in the area of data processing and also, concomitantly, the advances available to us in the fields of communication.

So we developed, with the cooperation of the advisory committee and the consultants from the System Development Corp. a feasibility study which was the beginnings of the creation of the New York State Identification and Intelligence System. That was in 1963, and in 1964 I had the honor of being appointed by Gov. Nelson A. Rockefeller as director of the emerging New York State Identification and Intelligence System, or NYSIIS, as we call it. My appointment was nonpolitical. I was never asked about my political persuasion, nor, indeed, has any member of the NYSIIS team been so asked.

The system was located in Albany, centrally located within the State of New York, and we dedicated ourselves to this concept of cooperative information sharing among criminal justice agencies, within a context of security and a climate of concern for the protection of individual rights and liberties. NYSIIS, therefore, was given no operational responsibilities and when a statute was passed officially creating the agency, it was established in the executive department, responsible directly to the Governor.

The bill which created NYSIIS and which I would like to later submit with your permission, Mr. Chairman, for the record, establishes NYSIIS as an independent agency having no powers, duties, or facilities to arrest, investigate, prosecute, confine, or supervise; no dual obligation, no feeling upon my part as director that perhaps I would be guilty of nonfeasance if some information indicated to me that I should pursue this investigatively.

It was conceived also, as a voluntary system and the participants were free to participate or not to participate, as they pleased, and to determine for themselves the degree of their participation.

Access to the information is to be security controlled, it is confidential and is not available to employers, to defense counsel, private detectives, or anyone other than the user agencies officially engaged in the administration of criminal justice. Nor, of course, would it be available to the mass media. Dissemination of information entrusted to the system is restricted in accordance with the wishes, the will, and the desire of those who contribute the information to NYSIIS.

NYSIIS restricts the input of information to avoid entry into the system and into the data base of wiretap information, such things as grand jury minutes, the identity of criminal informants, and likewise the system will not—

Mr. GALLAGHER. Excuse me, Director Gallati. You say "restricts to the system" or "restricts from the system"?

Mr. GALLATI. To bar from the system the identity of criminal informants, and likewise the system will not accept such information as tax information, social security, unemployment insurance, voting information, or family court data. These exclusions were the result of a long series of meetings of the NYSIIS advisory committee and this committee, as I mentioned before, represented leaders in all branches of criminal justice in the State of New York.

They made careful, considered, value judgments as to the inclusion and exclusion of various types of information which would or would not be included in the NYSIIS data base.

We were very much concerned from the outset about the problems of civil liberties and civil rights and constitutional guarantees. It is very difficult for anyone in the field of criminal justice today to ignore the imperatives of these considerations, particularly in the light of recent decisions of the Supreme Court; one has to be continuously alert to the implications of constitutional guarantees. And, of course, as in the *Miranda* decisions, we were reminded that we need to be more efficient in our criminal justice efforts and that we should utilize the facilities available to us presented through research and development, through science and technology, to do a better job in criminal justice.

Mr. ROSENTHAL. Inspector Gallati, does the statute that set you up limit the people that you can provide information to?

Mr. GALLATI. This sets up the limitation that it is designed for qualified agencies concerned with the administration of criminal justice. And it expressly states that this means courts of record, probation departments, sheriffs' offices, district attorneys' offices, State division of parole, New York City Parole Commission, State department of correction, New York City Department of Correction, and police forces and departments having responsibility for enforcement of the general criminal laws of the State.

Mr. ROSENTHAL. Can you give it to an outside State agency?

Mr. GALLATI. No, not to one who is not a user or who would come under these categories.

Mr. ROSENTHAL. In other words, you don't provide information if the State of New Jersey asks you for something about an alleged criminal, you would not provide the information?

Mr. GALLATI. This is not set up, yet, as an operating system, as I will mention shortly.

Mr. GALLAGHER. But with two friendly chiefs—a friendly chief in New Jersey and a friendly chief in New York—the friendly chief in New Jersey could call his friend in New York and get information that you have?

Mr. ROSENTHAL. In other words, they could short circuit the system.

Mr. GALLATI. Well, this is not now an operating system, Mr. Chairman, and the situation which you propose is one which we are considering in terms of our total development. We feel, in reference to this matter, that the improper utilization of the information in this system by any user would be one of those things which we would concern ourselves with greatly in terms of overall security and also in terms of the internal discipline of the users within the system. So that today, when sharing of information might conceivably be improper between two chiefs of police, a lesser penalty would derive therefrom in the eyes of the chief from breaching any confidentiality than would obtain in a system of this type where he would be cut off from sources of information because of his failure to comply with security directives.

Mr. GALLAGHER. Yes, but there might be a criminal in New Jersey that you have in your files and they are trying to cooperate with each other.

Mr. GALLATI. Well, today, of course—

Mr. GALLAGHER. In order to bring him to justice.

Mr. GALLATI. There is a series of cooperative endeavors among chiefs of police and people in the agencies of criminal justice which is on a voluntary cooperative basis between chief to chief, agency head to agency head, and it is this that we are trying to encourage in our own State so that we can maximize the information sharing which exists.

Mr. GALLAGHER. Therefore, this information could really be transferred to another State quite easily on an unofficial basis?

Mr. GALLATI. This information could not be transferred in the sense of the data base transferred to another State, except through interface techniques which would be possible if, for example, the State of New Jersey were to develop a system similar to this and we then considered the problems of interface.

Mr. GALLAGHER. No. What I am saying is you have a chief who is a user of your information in New York, properly authorized. Now, he acquires information and he gets a telephone call from somebody in New Jersey saying, "What do you have on John Jones?"

Mr. GALLATI. Well, of course, the user of the system will receive from the system that information which he has a need to know, a right to know, and which the contributor of the information has said this person can receive. Corruption, of course, is always possible; however, it will be extremely difficult and it will be heavily prosecuted and punished. I would like to remind the chairman that we did mention the fact that the person who contributes the information can put whatever restraints he desires upon the information which we then would be required to respect. In other words, if he were to say "I will put this information into the system but only my own agency can retrieve this" or "only my own agency plus this agency and that agency and that agency," we respect this type of constraint, and necessarily so.

Mr. ROSENTHAL. Does your computer have the technical ability to accept the information with restraint?

Mr. GALLATI. Yes. We are so advised by our computer experts, and I might state at this time that we have a man from System Development Corp. who is both a computer expert and an expert in the field of security, assigned full time to advising us on how we can develop the system along these lines.

Mr. ROSENTHAL. The answer to my question is yes?

Mr. GALLATI. Yes.

Mr. HORTON. One other question on this system—

Mr. GALLATI. I think Representative Horton wishes to ask a question.

Mr. HORTON. No; go ahead and answer his question first.

Mr. GALLATI. What I would like to point out is that the system is still in the developmental stage. The system will be operational in terms of building block 1 in August of 1967. Included in building block 1 on August 1967 will be, No. 1, the facsimile transmission system which we are now developing and are beginning to install. We will make primary installations or first installations in the early fall. We now have operating between Rochester and Albany and between Mineola, with the Nassau County police, and Albany, a testing device whereby we are able, through facsimile transmission of fingerprints, to transmit a set of prints from Rochester or Mineola to Albany within 14 minutes; a hard copy is received in Albany. It is there searched through our regular searching procedures and is returned—the criminal history response is returned again by facsimile within 4 minutes to Mineola or Rochester.

I would like to say—

Mr. HORTON. I would like to ask one question.

Mr. GALLATI. Surely.

Mr. HORTON. We have a quorum call and we are going to have to go over and answer the quorum. But are there any criminal restrictions on the users of this information concerning their dissemination of the information or is this wide open?

Mr. GALLATI. The general laws of the criminal code and the criminal law place restraints, of course, general restraints, upon the misuse of information, misfeasance of public officials and, of course, the failure of public officials to perform their duties in the sense of nonfeasance, as well. The general criminal law—

Mr. HORTON. In other words existing laws are considered to be adequate to cover the user's dissemination of this type of information?

Mr. GALLATI. As far as we now can anticipate, the answer would be "Yes," Congressman.

I point out, too, that we would also have the laws of bribery, corruption, laws aimed at corruption, and so on, which would be part of the total picture in terms of the discipline of the system from the user's standpoint.

Mr. GALLAGHER. Director Gallati, while I am thinking of it, you used the figure of 70 million. That is about one criminal for less than every third American.

Mr. GALLATI. Well, I think, if I may suggest this is more in terms, of the mass of data which is acquired by the number of autonomous agencies, independent agencies that are involved frequently with a limited number of persons. The point being, of course, that it does illustrate the duplication of files.

Mr. GALLAGHER. I see.

Mr. GALLATI. Many of these files are resting in one police department, also on the same person a file would be resting in another police department, and, likewise, as you go through the process of the criminal justice agencies; more files become built up as we go along. These files are being searched at the rate of 8 million files per year and they are being searched at great expense to the local communities.

Mr. GALLAGHER. Does a citizen have an opportunity to check to see if his name is in your file?

Mr. GALLATI. No.

Mr. GALLAGHER. As far as its accuracy is concerned?

Mr. GALLATI. No. We have no provision for that, Mr. Chairman.

Mr. GALLAGHER. Supposing Governor Rockefeller wanted to appoint somebody to a job, and a routine police check was made to see whether or not he has a police record, and your computer threw up the fact that there was a flag on him. Who determines the accuracy, or is it institutionalized for life that he might have taken a ride on the wrong bicycle at age 16?

Mr. GALLATI. The files, like criminal files generally, of course, are confidential. There would be in this particular instance you cite, where a public official would be investigated either by the Civil Service Commission, or, in the case of an appointed official, by the State police, they would utilize our files just precisely, in effect, as they now utilize the central identification files. Then it would be up to the appointing person or the appointing agency or the Civil Service Commission, as the case may be, to be guided by whatever rules they may have in terms of disclosure. Certainly, if any time that we are made aware of any data in the bank which might be in any way inaccurate, we would make every effort to remove it. I think that our files, potentially, would be considerably more accurate than those files which might be kept in local agencies where there would be less resources to keep them accurate and perhaps less resources to make sure that they are secure.

Mr. GALLAGHER. No. What I am saying is if a person is not on notice as to what is in his files, how would accuracy be checked?

Mr. GALLATI. Well, I would assume that when he has occasion to be accused of something or to—

Mr. GALLAGHER. He never is, he is just sort of a law-abiding citizen now, sufficient for the Governor or somebody to appoint to public office. It would now be very easy to ask for a profile on him from your computer. He doesn't know what is in there, but it is institutionalized—a rumor or a letter or whatever it might be that would cast an aspersion on his character. Therefore, the computer would throw out or not throw out his name as a bona fide risk.

Mr. GALLATI. Well, this would be something which would be within the purview of the executive chamber. We would respond with the type of information that we will have in the file, to the agency or the person or the official who makes the proper request. Our dealing in this case, of course, would be with the State police, with the Civil Service Commission, or an official user agency. So that we would then, of course, have a dialog between them and ourselves, if there were something that needed to be discussed in connection therewith. The real remedy necessarily lies in the judgment of the appointing authority and how he uses the information he receives—as it is today. NYSIIS

makes no such judgments, and indeed judgments will be made in the future by the same persons, in the same manner, as today.

I would like to point out that while we, of course, are concerned with the negative aspects of the possible violations of civil liberties, civil rights, and constitutional guarantees, that there are a number of very positive effects that this system has.

As a matter of fact, we are very much concerned with the positive effects just as we are with the possible negative implications. For example, we have had a very close relationship with the Vera Foundation and I would like, if I may, at the conclusion of my testimony to submit to you a letter from Mr. Sturz, the director of the Vera Foundation, in which he supports our efforts and recognizes the benefits that can be derived from being privy to, as soon as possible, the fact that a person does or does not have a criminal record which may relate, No. 1, to whether or not he should be bailed; No. 2, whether or not he should be summoned in lieu of arrest; and, thirdly, whether or not he should be released on his own recognizance at the time of judicial determination and also to what extent he should—

Mr. GALLAGHER. This recommendation is in the computer?

Mr. GALLATI. No, Mr. Chairman. May I clarify that, please? The point is that in the Vera Foundation studies on the bail problem and on the summons in lieu of arrest situation they found that it was possible, where a person could not otherwise raise bail, because of being indigent or not willing to take those steps necessary to notify friends, concerning the plight of the person arrested, that an investigation, rapidly conducted by representatives of the Vera Foundation, could determine whether the person had roots in the community, was not dangerous, not a dangerous person to the community, that he could be relied upon to appear for trial. This proved to be most successful and many people were released on their own recognizance on this basis. Likewise, many people, in experiments conducted by the New York Police Planning Bureau in New York City and by the Vera Foundation, it was found that a summons could be issued in lieu of arrest.

So that the indignity of a person spending a night in jail for a minor offense could be avoided, in cases, such as we found in the 14th precinct on 30th Street, where many housewives, people who were perfectly reputable, succumbed to a momentary decision to steal a slip or other item of clothing, and were arrested for shoplifting. These ladies because of their tremendous embarrassment were unwilling to notify their husbands or their parents, but were saved from this indignity of being locked up overnight with people of lesser repute. And, of course, the obvious saving that this ultimately will entail in terms of our unhappy police responsibility of keeping people in the lockup overnight, is quite obvious. But I would like to point out also that the Vera Foundation is now moving into the area of more serious crimes, crimes which are fingerprintable. The utilization of NYSIIS for the rapid transmission of fingerprints in those cases where fingerprints are available, of course, means that many people will not have to be kept in jail overnight, and in the alternative, may be summoned and, in other cases, released on their own recognizance.

Mr. GALLAGHER. Are you tied in with the Federal Bureau of Investigation?

Mr. GALLATI. No, we are not. The records in the State of New York have been utilized for these purposes by the Vera Foundation and have been found to be—

Mr. GALLAGHER. No. I meant on the transmission of fingerprints.

Mr. GALLATI. No, sir, we are not.

I would like to also mention in this connection that the judge has, through this type of system, at his fingertips a great deal more information than he would otherwise be able to assemble quickly in reference to the individual whom he must deal with. And, therefore, he can more discreetly or in discreet fashion sentence in those cases where a conviction results or may release on his own recognizance the person before him, if this is an appropriate decision to make. If the court is uninformed, the court is liable to err on the side of keeping the person in jail.

I would like to continue the point which we made about the process of criminal justice at the point of arraignment, because I think here again we have a situation where, through the utilization of this facsimile system, which is one of the parts of Building Block One of the system, we will be able to overcome some of the very problemful areas of rapid arraignment requirements. I am sure we are all aware of the need for rapid arraignment and the implications that are bound to arise in those cases where arraignment is unduly delayed.

Now, one of the problems that the police have always had has been this problem of making sure what the prisoner represents and, of course, fingerprinting gives us complete, thorough and incontrovertible identification if we have a set of prints for this person on file in the Central Identification Bureau. And also, of course, a record of his criminal history, if such is the case.

Now, the situation in which the police are placed in terms of rapid arraignment is this: When they have a prisoner and they can legally fingerprint the prisoner, they must learn one of two things. They must either assure themselves from the criminal record received, as the result of the submission of the fingerprints, that this person is not wanted for a more serious crime elsewhere, or, is in fact wanted for such a crime and must so advise the judge, upon arraignment, as to the status and the criminal record of the person whom they present to the arraigning magistrate.

The seriousness of this is not to be underestimated. For example, in the city of New York the police department maintains at considerable expense to the city of New York a bureau of identification which is very largely redundant and duplicates the same type of file, although perhaps more extensive, maintained by the State in Albany, but, because of the requirements of the New York City criminal courts and the desire of the New York City Police Department to present what is known as the yellow sheet the criminal history at the time of arraignment before the magistrate, they have maintained in the city of New York at a considerable expense, perhaps in the neighborhood of \$2 million yearly, a duplicate type of facility to enable them to respond in that fashion.

We feel with this new facsimile system and our increased capabilities for the search of our fingerprint files that we will be able to respond in such fashion that rapid arraignment will be possible in New York City based upon those files which we in NYSIIS maintain, but

also similar rapid arraignment will be available to, and discreet arraignment, I might add, available to the chief of police or the detective in any department, regardless of how small, throughout the State.

So, we see we have the potential for civil liberties advances in the arraignment aspects, in the summons aspects, and also in terms of bail and sentencing.

We also, Mr. Chairman, if I may ask you to accept this letter, have a letter which we have received from the New York Civil Liberties Union likewise supporting the aims and goals of NYSIIS.

I would like at the conclusion of my testimony to ask you to consider that for the record.

Mr. GALLAGHER. Very well.

Mr. GALLATI. I also would like at this time to call your attention to a brochure which we have issued and distributed throughout New York State to people in the area of criminal justice, professionals in this field, and to many members of the public entitled "Information Sharing, the Hidden Challenge in Criminal Justice." This describes, I believe, rather well the goals, the aims, and the directions taken by the New York State Identification and Intelligence System. I would ask if you would like to accept this at the conclusion of my testimony, Mr. Chairman. (See p. 159.)

Mr. GALLAGHER. Very well.

Mr. GALLATI. We also have received support from the Association of the Bar of the City of New York, the law enforcement committee thereof, and also the committee on the criminal courts the records of which I would like at a later time to submit to the subcommittee.

We feel that this information sharing capability should be made available to the individual units, which I might add are very numerous, there being 3,600 agencies—separate agencies—of criminal justice in the State of New York, 611 of the 3,600 are police agencies, be they sheriffs, local police departments, city police departments, county police departments, State police, or otherwise—

Mr. GALLAGHER. Do you have machines tied into a network with all of these agencies?

Mr. GALLATI. We will have within the next budget year a total of 40 installations throughout the State of facsimile devices which will provide sending and receiving equipment.

Mr. GALLAGHER. This is a telecommunications system.

Mr. GALLATI. It is a facsimile device attached to normal communications lines, that is telephone or Western Union. We are in the process of evaluating these lines.

Mr. GALLAGHER. What protection do you have against interception?

Mr. GALLATI. We have anticipated the problems of the communications systems, and we are looking very seriously at the devices available such as those used by the military for scrambling and so on, which will permit us to prevent any kind of electronic interception of these transmissions.

Mr. GALLAGHER. Is there such a device now available?

Mr. GALLATI. It is my understanding that there are such devices and that they can be applied to our transmissions.

Mr. GALLAGHER. Is that your understanding, Mr. Baran?

Mr. BARAN. Military cryptographic units are extremely expensive. With the large number of terminals they are eventually talking about, cost could be an overwhelming factor.

Mr. GALLAGHER. What would be your estimate of cost for such a transmission system fully safeguarded against interception?

Mr. GALLATI. We are not able at this time to give you any final costs on that, Mr. Chairman.

Mr. GALLAGHER. Could you give me an estimate for a 40-unit system?

Mr. GALLATI. I can give you estimates of the cost of the units but not of the scrambler, sir.

Mr. GALLAGHER. It is my understanding—and I stand to be corrected—that while a telecommunications system is a relatively simple or inexpensive system to set up, when you start programming, or constructing safeguards the system now starts to multiply from 3 to 10 times. Is that your understanding?

Mr. GALLATI. No; I have not had that understanding exactly.

Mr. GALLAGHER. So we have the benefit of wise counsel, Mr. Baran, is that reasonable?

Mr. BARAN. That seems reasonable. That order of magnitude.

Mr. GALLAGHER. You do not accept that or you do not know who he is?

Mr. LUMBARD. That is right. We do not know who he is or whether he is right.

Mr. GALLAGHER. I know who he is and I know that he has done considerable work on this subject at the Rand Corp., which does some of the big thinking for some of the major problems that confront the military and our Defense Establishment. So he does have some expertise in this department. That is why if you had some information on cost analysis we would be very interested. In fact, I think the Defense Department would find it helpful.

Mr. GALLATI. I do not have it at this time, Mr. Chairman.

Mr. GALLAGHER. This is the key to the whole problem we are faced with.

Mr. GALLATI. I think what I should point out here—and perhaps I am derelict in not having pointed it out before—is that we do have a hierarchy of confidentiality within the system. The Building Block One of which I now speak, which we primarily desire the facsimile devices to support, is a relatively low secrecy type of operation. Perhaps I best might run through the types of things we anticipate having in Building Block One. Let me explain also that it is our concept of development of this system which has been developed at a very deliberate pace, to search out the correct course very carefully through systems analysis procedures and considerable study each step of the way. It is a building block approach and a modular amplification type of building block development. For example, as I said, in Building Block One we will have this facsimile system. The facsimile system is tied in with the other aspects of Building Block One, such as the more rapid response of criminal histories to a fingerprint request, fingerprints having been, in most cases where rapid response is sought, taken as the result of an arrest situation. So we are improving the inhouse turn-around time of the DCI which is now in terms of several days and hopefully it will very shortly be reduced to 1 or 2 days and then to a period of 1 or 2 hours. This is the second module, if you will, of Building Block One.



The third module would be the development of a fraudulent check system. One of the problems of law enforcement in New York State, as I presume it is elsewhere, is the problem of fraudulent checks. Here again, the facsimile system ties in well with the transmission of the fraudulent check from the individual localities throughout the State to a central location where it can be checked to see if the person who has written the check has forged this on other checks and is possible of identification.

Mr. GALLAGHER. How do you identify the caller for the information?

Mr. GALLATI. The callers will have a special code number. When we get into more sophisticated and more detailed security problems, when we go down the road through this hierarchy I speak of in terms of degrees of security, the rank order of security, we will have other devices which we anticipate considering for this purpose. The State of New York is developing a network of closed-circuit television and we feel this will be very helpful to us in terms of our utilization of security measures. The closed-circuit television will permit us to have almost the type of thing that we now have and which so many people and the police feel is so essential; namely, face-to-face confrontation when you exchange information.

Mr. GALLAGHER. You do not feel this can be intercepted very easily?

Mr. GALLATI. Excuse me, Mr. Chairman, I missed the point.

Mr. GALLAGHER. My point is: How do you identify a user or a caller as a bona fide caller?

Mr. GALLATI. No. 1, the agency would be on the system and would have control over the utilization of that particular input-output device which is located in official places such as police stations or courts or district attorney's offices.

Mr. GALLAGHER. Anybody with a proper signal in his cellar might be able to call that information and put it on his wide screen. This is one of the problems that disturbs us. This is why I would like to know your experience.

Mr. GALLATI. We have a coding system which would have to be known by the person who makes the inquiry. He would have a coded inquiry which would then permit him to receive the type of information which he is entitled to receive. The rank order of security, and the hierarchy of the types of things which we will have to provide security for, will begin with limited need for sophisticated security devices. For example, the criminal history sheet which is now available to be transmitted over facsimile machines, or perhaps the fraudulent check which would trigger a criminal history response if the person is identified; this type of information, while it is confidential, is of a relatively low order of confidentiality. It would not be likely that any tremendously sophisticated operations would be involved in trying to intercept this type of information, because it would probably not justify any tremendous involvement of persons who would try to interfere with this. However, we are not overlooking the need for maintaining security both at the point of entry into the system, at the point of reception of the information, personnel security, communication security, physical security, and, of course, actual line security. Again, we are considering this and we will give to each module as it develops the type of security which we feel it deserves in terms of the sensitivity of the material itself and the threat that we have to expect.

Mr. GALLAGHER. Do you have any cost estimates at all on this sort of thing?

Mr. GALLATI. The scrambler which we have looked into was relatively simple and was not tremendously expensive. I do not have the exact figures available at the moment. I would like to make the distinction between the type of scramblers that would be required for very highly sensitive information which could possibly be compromised and which would be of such sensitivity that the threats to it would be so great that it would require these types of tremendous protective devices, as opposed to a much lesser type of protective device which we feel in the Building Block One context, with the type of modules we have on Building Block One, would be adequate for the purposes. I might mention some of the other factors which we anticipate developing as modules for Building Block One. I have spoken of the facsimile system, the development of a more rapid response situation within the division of identification. I have spoken of the fraudulent check module. I would like also to speak of the development of a single fingerprint module, which gives me an opportunity to mention that we have considered in this system the problem of squeezing geography so we have information sharing despite the miles between independent agencies, and we have spoken also of the vertical dimension in terms of the various agencies of criminal justice which all have similar goals and operate in reference to the same types of persons and the same individuals in many cases. I would like to mention also the third dimension, which is one unique to our development, in our opinion; that is, the application of science and technology to developing better operational techniques. We are well aware of the famous computer caveat, GIGO or garbage in and garbage out. We did a system analysis of the problem of latent or scene of the crime fingerprints and we found that these were much less effective than they should be. The public has been lulled into a false sense of security in terms of the utilization of scene of the crime prints by law enforcement people.

Mr. GALLAGHER. This was not the same crooked computer you had in New York, was it, that was used to sell passing marks to police?

Mr. LUMBARD. That is news to me.

Mr. GALLAGHER. I read a story about a person taking a civil service examination and a friend stood next to the computer and put in the right answers. It is one of the weaknesses of the system that I sometimes find. Not only can you take bad information out but you can put the wrong information in.

Mr. LUMBARD. That happened in New York?

Mr. GALLAGHER. It was in the New York Daily News.

Mr. LUMBARD. About a computer in the New York State Civil Service System?

Mr. GALLAGHER. About a computer that was grading civil service examinations.

Mr. LUMBARD. In New York State?

Mr. GALLAGHER. Yes.

Mr. LUMBARD. I would very much appreciate that clipping because that is news to me.

Mr. GALLAGHER. It was news, the Daily News.

Mr. GALLATI. I think we can all agree if you put garbage into a computer you will get garbage out. You do not get anything more

out of a computer in most cases than you put in. Therefore, to merely put more information in a computer and take an ineffective operational technique and magnify that information which is utilized in the operational technique is in a sense multiplying fractions of effectiveness. So we have done a very, very considerable amount of research and study on the problem of scene-of-the-crime prints.

I feel that this is appropriate to call to your attention because I think it does relate to what a computer can do. The Henry system of fingerprinting is now over 60 years of age. It was obviously conceived of and developed long before the age of computers. It is not computer-compatible and it has never served us well as in terms of the scene-of-the-crime prints despite what one might read in detective novels. In order to search a large fingerprint file it is necessary to have a completely visible and readable set of 10 prints from the individual. It is possible in the great bank robbery in England, or the Weinberg kidnapping, to make special efforts to go into your main files. But barring this, the normal procedure for a scene-of-the-crime print is to search it through what is known as a single fingerprint type of system. In New York State where we have I am advised is the largest single fingerprint file of any State in the Nation, we have only 20,000 people in that latent print comparison file, 200,000 individual prints. So if you do not happen to be, as a burglar or a car thief, in that particular file, your chances of being identified by having left a chance impression, anything less than a conveniently rolled set of 10 prints, means that you are "home free."

Only twice in the history of the State of New York has any criminal been so cooperative as to leave us a perfect set of 10 prints.

We have, of course, a very large file in New York State. If we can get into this total file of millions of prints as opposed to 20,000 prints, we are that much better off. I believe that the Federal file is no larger than that of New York State in terms of single fingerprints. So we see the same type of problem at the Federal level. In other words, you have this tremendous number of sets of prints which are not normally searchable and retrievable by any scene-of-the-crime print process, but only by the careful rolling and ink printing of individuals, that is, the normal way in which to assess these files.

Mr. GALLAGHER. Mr. Director, I think I sense an instantaneous retrieval call on me to get over to the House and vote. I think I am going to have to leave. I was hoping that Mr. Horton or Mr. Rosenthal would be back. I would like to say that we appreciate very much your coming down here and giving us the benefit of your experience. We will be happy to include in the record the documents that you have offered.

(The documents follow:)

NEW YORK CIVIL LIBERTIES UNION,  
New York, N.Y., March 18, 1965.

Mr. ROBERT R. J. GALLATI,  
Director, Identification and Intelligence Project,  
Albany, N.Y.

DEAR BOB: Thank you for forwarding copies of senate introduction 1481 and assembly introduction 2625 creating the New York State identification and intelligence system in the executive department. As you know, I have been much impressed by your efforts to improve the processes of criminal justice for many years, and I have been most pleased with the emphasis upon civil rights and liberties which you stressed both as dean of the police academy and in your

field operations. Your continued concern for enhancing civil liberties through computer-based information sharing among the agencies concerned with the administration of criminal justice, in my opinion, is further evidence of your sensitivity to the protection of individual rights.

I fully agree with you that accurate, informed, official decision at every stage of the process of criminal justice will help to safeguard the civil liberties of suspects. Unnecessary arrests and futile commitments can be avoided where adequate data concerning the individual is rapidly available. To the extent that law enforcement becomes more scientific and utilizes sophisticated technology such as you are developing, there will be continually less justification for excessive interrogation and detention prior to arraignment. Where investigative leads and evidentiary material can be supplied or developed through the rapid retrieval of discrete data, it should profoundly improve traditional police practices in the entire area of arrest, search, and seizure.

I wish you well in your sustained efforts to achieve higher levels of professional practice in the administration of criminal justice. On behalf of the Civil Liberties Union let me say that your objectives are consistent with our own and we are delighted to be kept advised of the results of your research.

With best wishes for your continued success,

Yours,

GEORGE E. RUNDQUIST,  
Executive Director.

THE VERA FOUNDATION, INC.,  
March 16, 1965.

Mr. ROBERT R. J. GALLATI,  
Director, Identification and Intelligence Project,  
Albany, N.Y.

DEAR BOB: Many thanks for sending along the brochure entitled "Information Sharing: The Hidden Challenge in Criminal Justice." It is a highly professional job and should serve to interest a great many people in the benefits to be derived from knowing as much as possible about persons suspected of violating the criminal law.

As you know, Vera Foundation bases its recommendations for summons in lieu of arrest and the release of indigent arrestees on their own recognition on a rapid research of the individual suspect's reliability. We see the system as a valuable aid in the operational extension of the concepts developed in the Manhattan summons project and the Manhattan bail project.

We are now planning to work in the area of arrests of alcoholics to see if more humane methods of dealing with this problem might prove feasible. Here again, the more one knows about the person with whom he is dealing, the more rational can be the treatment of the individual and the better human rights can be safeguarded.

On behalf of Vera Foundation I would like to assure you that we support your endeavors and endorse the New York State identification and intelligence system objectives so ably presented in your brochure.

Kindest personal regards.

Sincerely,

HERBERT STURZ.

# THE NEW YORK STATE IDENTIFICATION AND INTELLIGENCE SYSTEM

## INTRODUCTION

### The background

As in other areas of the Nation and the world, New York State is in the first stages of a monumental population explosion. In 1960, the population of New York was 17 million. By the year 2000 it will have almost doubled—estimates are at 30 million.

The same expansion that is filling our schools and expanding our economy will, unfortunately, tax our system of criminal justice and correctional institutions. Increased population mobility creates different problems—especially coordination. Today, the average American family moves once in 5 years, or one-fifth of our families move every year.

Some statistics emphasize the effect of the population explosion on the administration of criminal justice. In 1962, in the State of New York:

93,000 lawbreakers were either in direct custody or under legal restraint through probation or parole at any one time.

3,636 agencies concerned with the enforcement of criminal justice were employing more than 66,000 people at an annual cost of \$665 million to local and State taxpayers

3 million alleged violations passed through the courts of New York, and 700,000 felonies, misdemeanors, and lesser offenses (excluding traffic violations) were reported to the police.

Predictions for the future provide no cause for complacency:

By 1965, 50 percent of the State population will be 25 years of age or under; and statistical studies have shown that the high-violation group consists of persons between the ages of 16 and 25

Crimes reported are increasing; e.g., 2 years ago, the national crime rate was increasing at four times the growth rate of the population; today, this national rate has risen to five times the population growth.

Fortunately, New York State is somewhat below these national averages, but clearly, there is an urgent social need to combat crime more effectively and efficiently. New tools and fresh efforts are needed; one possibility is for agencies concerned with criminal justice to adapt and employ the resources offered by modern technology in communications and electronic data processing. Just as the general population is more mobile, so is the criminal; criminal activities no longer are confined to a single community or localized area. It is difficult, if not impossible, for local agencies to maintain all files necessary and adequate for their needs. Yet one of the most vital tools in the administration of criminal justice is information—information that may be used as a source of leads in the investigation of crimes, as the basis for preparing criminal charges, as evidence in a court of law, as background for bail and sentencing, and for rehabilitation programs. Under our system of law, information must be utilized at all stages of the process to develop action.

#### *The information problem*

Although agencies concerned with the administration of criminal justice have the common goal of protecting the person and property of our citizens, thus maintaining law and order, under the existing structure, each has a separate, manually maintained information system that evolved to suit the purposes and problems of an earlier age. These information files contain many duplications; they vary in scope of content, coverage of individual subjects, format, and currency. When agencies borrow data from each other, the copies and extracts further duplicate and swell the files with a consequent rise in clerical costs and storage space.

The current number of forms in the activity files of State and local agencies concerned with the administration of criminal justice is estimated at 60 million; in 1971, it is estimated that this figure will rise to 70 million. More than 8 million searches of these files are performed annually, and as the crime rate increases with an expanded population, so will the required number of searches.

The State of New York under the leadership of Gov. Nelson A. Rockefeller has taken many effective measures in the war against crime—among the more recent, the establishment of the New York State Identification and Intelligence System (NYSIIS). The concept for this system had its genesis in the investigation of the notorious "Apalachin meeting" in November 1957 of more than 75 of the Nation's top criminals. Even a summary account of this investigation dramatizes the nature of the information problem the system was conceived to solve.

Following the discovery of the Apalachin convention of criminals, the New York State Commission of Investigation was formed in 1958 to undertake an all-out inquiry into the meeting—to identify its attendees and to determine its purpose. Aside from the silence of the meeting's participants, the mechanical tasks of assembling information on these major criminals proved to be formidable: just one of them was the subject of as many as 200 separate official police files in a surrounding area of several hundred miles. Two years later all of the State's files on these criminals still had not been accumulated or assimilated. What the taxpayers had purchased was not brought to bear upon a pressing problem. Those files that were examined often were barren of original material—a relatively meaningless collection of newspaper clippings, copies of other files, and loose notes. The distillation of valuable content was minute.

As chief counsel of the commission, Eliot H. Lumbard was intimately concerned with that investigation and its frustrations. When he was appointed to

Governor Rockefeller's staff in 1961 as special assistant counsel for law enforcement, among the assignments he undertook was a search for possible remedies to the problems presented by the Apalachin investigation, as well as the deluge of papers and files inundating operational agencies concerned with the administration of criminal justice. Out of his inquiries into the possible application of electronic data processing techniques to criminal justice came the New York State Identification and Intelligence System project.

In the last 10 years, the development of new electronic data processing equipment—particularly the electronic computer—has stimulated revolutionary changes in information handling techniques in many fields. Computer-based information processing systems have been developed for military (e.g., air defense) and commercial (e.g., airline reservations) purposes. These computer-based systems make possible the centralized and organized storage of large amounts of data that can then be drawn upon by a network of geographically dispersed user-agencies equipped with communication devices and provide the users access to this central store of information with a speed, accuracy, and unfailing memory unrivaled by any manual methods.

#### *People and plans*

In May 1963, Governor Rockefeller, with the concurrence of the legislature authorized a study to determine the technical feasibility and potential significance of applying this new electronic technology to the information problems of the New York State agencies concerned with the administration of criminal justice.

Initial system analysis was performed by the System Development Corp. During this study phase, guidance was provided by the executive chamber and, under the direction of Donald Axelrod, the administrative management unit of the division of the budget of New York State. Throughout the feasibility study, the project derived considerable benefit from a broadly-based advisory committee that met frequently to provide information and advice. Advisory committee members were representatives of—

- The executive chamber.
- The State department of correction, including probation representatives.
- The State division of parole.
- The division of State police.
- The division of budget.
- The judicial conference (the administrative arm of the courts of New York).
- The New York State Association of Chiefs of Police.
- The New York State Sheriffs Association.
- The New York State District Attorneys Association.
- The Office of Probation for the Courts of New York City, and
- The New York City Police Department.

A feasibility report, dated November 1963, resulted from that 6-month study and concluded that a central information processing system was technically feasible and that it could provide valuable assistance to all the relevant agencies.

Governor Rockefeller, in his annual message to the New York State Legislature in January 1964, urged an appropriation to initiate the development and installation of the proposed system. "The increase in the incidence of crime, the modern techniques of organized crime syndicates, and the increased mobility of our population make it imperative that law enforcement officials use the most modern technology in collecting, storing, retrieving, and disseminating information and intelligence and engage in more effective cooperative action in the area of information sharing."

New York State, operating under the system of annual budgets, provided funds for the first year of development. Robert R. J. Gallati, formerly chief of planning of the New York City Police Department, was appointed director of the project. To assist Gallati, Paul T. Veillette, formerly director of the management analysis division of the Chicago Police Department, was appointed technical director, and Paul McCann, formerly director of the division of identification of the New York State Department of Correction, was appointed technical consultant. A growing State staff supports them. The executive chamber, the administrative management unit of the State budget division, and the advisory committee continued to provide support and guidance to the project.

System Development Corp. was selected by New York State to assist with development of the system. A group teamwork approach has been adopted by

New York State and SDC for unified project planning and joint execution of the project.

#### THE PLANNED SYSTEM

##### *The basic concepts*

The New York State Identification and Intelligence System, the first of its kind in the world, has as its long-range goal a computer-based central facility, located at the capital city of Albany, that will store all the information pooled in the system. Qualified agencies concerned with the administration of criminal justice, located anywhere in the State, may contribute and have access to this information center on a voluntary basis via a communication network.

The system's approach is based on the concept of information sharing among all agencies involved in administering criminal justice; these agencies fall into six broad functional groups—police, prosecutors, criminal courts, probation, correction, and parole. These agencies are necessarily independent and autonomous in their operations, yet none of these agencies provides a service that is a totality of and by itself; none functions alone. Each of these agencies has a part in a continuous operation; one subject, and his accompanying file, for example, go successively through these respective agencies until the ultimate disposition of the particular case. With rare exceptions, none of these agencies, except the police, may take official action unless it is on a matter in which another agency has previously acted, and then it is usually concerned with the predecessor agency's "product." Each agency needs to have the same information, with some variations, about each subject, yet each is currently served by separate information sources.

At present, the sharing of information among agencies is uneven in quality; i.e., the transfer of facts often is unsystematic and undependable. A major problem has been the difficulty of developing regular procedures for movement of information among agencies that are involved in different aspects of public service, but are in possession of information of great value were a means for dispensing it to other agencies available.

For example, how is fact "A" in the official files in Buffalo to be correlated with fact "B" in Albany and, perhaps, additional facts in New York City? At present they are not coordinated. Upon full implementation, the new system will provide unified and complete information to all appropriate agencies without compromising the sensitive nature of the information; it will not inhibit their independence of operation, but will provide unity—of information—where unity is urgently needed.

While the system will represent a new agency in State government, independent of all present agencies, it will have no operational powers (such as arrest, prosecution, confinement); it will be strictly a State service agency and will not diminish the authority or autonomy of operation of any of the participating agencies. The system is not a fact-gathering device; it merely makes facts already being gathered by these agencies at taxpayer expense significantly more useful. It is a service tool for these agencies. Its routine functions can be operated effectively by present personnel of the user agencies; the system is basically simple and does not require the creation of an entirely new professional staff or new physical facilities.

There are 3,636 separate and virtually independent agencies that are potential State and local subscribers to the system. Among these are:

- Six hundred and eleven police agencies.

- Sixty-two district attorneys.

- Seventy-five probation agencies.

- Two thousand four hundred and seventy-nine courts with criminal interest, including 2,278 justices of the peace.

- Four hundred and six institutions for custody and detention.

- Three parole agencies.

Selective use of the system can be made by many other agencies, including:

- The Waterfront Commission of New York Harbor.

- The New York State Department of Motor Vehicles.

- The New York State Department of Law.

- The New York State Division of Alcoholic Beverage Control.

- The New York State Department of Mental Hygiene.

##### *System capabilities*

In support of the normal daily operation of these participating agencies, the fully implemented system will provide rapid access to summary criminal history, as well as detailed criminal, social, and modus operandi data on each subject;

will rapidly transmit graphic data, such as photographs and fingerprints, throughout the State; will maintain files of personal appearance data, latent fingerprints, fraudulent checks, warrant-and-wanted notices, stolen motor vehicles, stolen property, laundry marks, stocks and auto registration forgeries; and will provide direct scanning and computer-based searching of all fingerprints on file, arrest and disposition reports, and intelligence information. All of these procedures will be performed with greatly increased scope, accuracy, and efficiency.

The function of fact coordination, with speed, accuracy, and completeness, is the principal contribution of the system in the battle against crime. For example, many advantages are derived from the data files in 1 police agency being immediately available to the other 610 police agencies of the State and to additional agencies performing subsequent activities in the administration of criminal justice. All agencies enrolled in the system, including district attorneys and probation officers, will provide certain types of data in their possession that pertain to their activities and gain direct access to appropriate identification, criminal history, and intelligence files on cases of interest. Courts, correctional institutions, and parole officers may also contribute facts, thus constantly enriching the data base and, in return, make requests of the system for additional necessary information.

However, all types of information would not—in fact, should not—be placed in the central information pool. The advisory committee felt all family court subjects and information should be excluded. District attorneys surely would not place information held in the secrecy of a grand jury proceeding in the system. It is not likely that data such as wiretap information, names of confidential informants, or unverified tips and rumors would be included. The omission of this information, totaling perhaps a fraction of 1 percent of the potential data, in no way affects the value of the entire system.

Data on motor vehicles (except for stolen automobiles and cars owned by major subjects) will not be included in the system because of potential duplication of the data files of the Department of Motor Vehicles system; that material will be readily available through what is called an interface between the two systems. The system will not attempt to duplicate all of the detailed information currently stored in the individual case jacket and folder-files of participating agencies. These files contain considerable detail on particular events, such as indictments, statements of witnesses, daily logs, or detailed descriptions of routine contacts, and parole and probation violations. Such information can be obtained, when it is needed, directly from the appropriate agency's files. However, the system will include adequate references to those sources, such as indications that a subject is currently on probation or parole or is confined in a particular facility.

The installation and operation of the New York State system will make it possible to reduce much of the present duplication of effort among agencies. Because of the duplication of effort, for instance, problems now occur in maintaining current files and in cross-referencing data in different files. These problems can be markedly decreased by an effective concentration of the multiple files into the single file structure of the new system. The system provides, also, for the constant enrichment of the file through mechanisms for updating, deleting, or adding new information. This efficient and appropriate storing and maintaining of data also helps to reduce costs by releasing personnel, space, and resources for other necessary duties.

In summary, the ultimate system as envisaged would provide four major new capabilities. First, suspects from files of known criminals can be identified to a degree and at a speed that has hitherto been impossible. Second, outstanding open cases can be solved by automatic and comprehensive searches that can be conducted on the basis of statewide information. Third, automatic abstracting, indexing, and retrieval of textual intelligence and modus operandi information can be used to obtain a comprehensive search of all documents stored in the system and a retrieval of only those documents that are relevant to the interests of the user, instead of the current time-consuming manual index card files and bulky folder files now provided.

Finally, the system offers the tremendous potential of being able to perform research: a pattern analysis of data to assist in evaluating new trends in crime, in testing new approaches to the administration of criminal justice, and in discovering patterns of structure and activity of criminal and special organizations in order to discover relationships in vast bodies of facts that would be almost impossible to develop through standard manual techniques of file searching.



*Safeguards*

Security precautions will be taken to maintain both the internal integrity of the system and freedom of use by the participating agencies. Interchanges of facts within the system will be made to users on a confidential basis within specific legal, right-to-know and need-to-know requirements. Users will be restricted to a defined category of public officials and agencies. Information will not be available to the general public, subjects of investigation, news media, or private attorneys. If so requested, selective, highly sensitive data will be tagged in the system to insure that its distribution is under the exclusive control of the contributing agency.

Following the commission of a crime, the investigating agency assembles various forms of investigative leads and potential evidence—physical or otherwise—such as testimony. These include latent fingerprints, intelligence data, photographs, handwriting samples, and descriptions of property left at the scene of the crime.

In addition, it attempts to furnish descriptions of the manner in which the crime is committed and, if possible, a description of the alleged criminal. This material, transmitted to the central facility, is compared to relevant files that are stored either in the computer or on microfilm.

If any of these comparisons are successful in providing a match, the system supplies the investigating agency with a list of possible suspects. Upon request, it will provide further specific information that may be stored in this system concerning these suspects.

In addition, the system maintains continually updated information on open cases and stolen property. When the investigating agency arrests a suspect, and if the crime is a felony, the agency prepares an arrest report and a fingerprint card, which are transmitted to the central facility by means of a facsimile transmission device.

There, names and prints are searched for a match with relevant files. If either search is successful, the criminal record is retrieved from the computer and sent to the arresting agency. This agency takes the criminal record together with available intelligence information along with the suspect into court for arraignment and determination of bail.

In the meantime, the system notifies the judicial conference, the court monitoring agency in the State, of the arrest so that the conference can keep track of the court transactions with its own data processing system.

The district attorney begins to prepare his case. He may ask the system for additional relevant information on the crime, including a list of unsolved crimes that resemble the one being prosecuted. If any of the unsolved crimes listed by the computer can be associated with the suspect, they may also become the subject of a prosecution.

If conviction results, the court may ask the probation officer to make a presentence investigation. He may begin his work by obtaining background data on the defendant and on his crime from the system.

If the defendant is sentenced to an institution or put on probation, the appropriate agency registers its custody and any significant events happening during custody with the system. If the offender is placed under parole supervision, a similar process takes place. In this way, the system keeps track of the individual as he moves from one custodian to another and stores this record in the event of any future contact with the State. The data base of the system is thus continually enriched. In summary, a variety of users, each concerned with a particular aspect of the administration of criminal justice, are served by the system.

This system represents a considerable advance in the coordination of relevant facts that, up to now, have not been coordinated and, as a result, are not readily accessible in the complex interrelationships among agencies involved.

The objective of the system is to collect, maintain, and coordinate the maximum amount of relevant data for New York State and local user agencies.

The concept is to provide, through a data processing system, a central store of accurate, complete, and timely information while reducing redundancy and mass of the data and bringing about long-range financial benefits.

## SYSTEM IMPLEMENTATION

*Building blocks approach*

In order to implement the system, a "building blocks" approach was evolved: particular sets of operational capabilities (including the techniques and data) will be produced for the system in modules or blocks; at the same time, system

analysis will proceed on future "blocks" of the total system. This parallel action assures the State of an early operational capability, followed by planned, orderly growth of the system that continuously provides increasing support to the using agencies.

The parallel action plan also permits the system to meet new needs as they arise as a result of changes in the law, user agencies, communities served by the agencies, and in available techniques and ideas. Presently, not all crimes require that fingerprints be taken; as the law changes or additional crimes require fingerprints, the system must be prepared to accept the increased load. From time to time, changes in the community require local agencies to perform new or augmented functions; these services, as well as the application of new techniques to the administration of criminal justice, may require additional support from the system.

The approach will utilize a number of long-term research endeavors that come to fruition at different times depending on the complexity of the problems involved; e.g., the system will be ready to utilize direct scanning and classification of fingerprints when it becomes feasible. Through incremental development, each new capability can be integrated with the existing system as the required technology is developed. This organic development focuses new research to support the administration of criminal justice.

The first block of the system is designed to provide capabilities that can only be gained from the sophisticated services of a computer. It will aid in the solution of crimes by providing rapid access to summary criminal histories; the computer complex will supply more information than is currently available from any one agency. Improved communication facilities, including facsimile devices, will rapidly transmit graphic data, such as photographs and fingerprints, throughout the State.

*Civil liberties advances*

This speed of communication will also realize advances in the protection of civil liberties. Accurate, informed decision by the judiciary will be possible when setting bail or releasing arrestees on their own recognizance; disposition and sentencing, especially of lesser crimes and offenses, will similarly be expedited.

Building block No. 1 will computerize a personal appearance file, a latent fingerprint file, a file of fraudulent checks, and storage and searching of arrest and disposition reports. These capabilities were selected for implementation on a computer on the basis of a system analysis that showed that computerized versions of these files would provide major improvements over the manual techniques in use today. During building block No. 1, the file of fingerprint cards and the name file of the division of identification of the State department of correction will provide major sources of data for the system on a manual basis. Subsequent research and development will lead to computerization of these files with a significant increase in service and efficiency.

Facsimile transmission equipment utilizing present communications facilities will provide user access to the central facility. This network will permit the rapid transmission of textual reports, fingerprints, handwriting samples, maps, photographs, and other forms of data back and forth between the user agencies and the central facility. The goal of building block No. 1 is for the equipment to provide the central facility with sufficiently clear copies of fingerprints to permit classification and matching.

This initial increment of the New York State Identification and Intelligence System will materially aid all the various State and local police agencies concerned with administering criminal justice. A print found at the scene of a crime in a small community could be checked within a short time against the latent fingerprint file in the central facility; a small agency would not likely have an extensive enough fingerprint file. The files of the entire State, then, would be at the disposal of any local community in which a crime was committed.

Similarly, a physical description of a suspect could be compared against data in the central facility; the rapidity and completeness of computer search increases the probability that both large and small agencies could identify individuals of interest to law enforcement and other agencies administering criminal justice. The system will make possible the communication of an up-to-the-minute summary criminal history to any governmental agency concerned with the administration of criminal justice. The fraudulent check capability should materially aid in the investigation and apprehension of individuals passing fraudulent checks as a result of the speed and thoroughness of investigation made possible by the use of electronic data processing.



Present plans call for the system to become operational with building block No. 1 by mid-1967.

Building blocks 2 and later: Although the full range of capabilities has not yet been determined—special studies are required before specific modules can be selected—there are a number of tentative modules that are likely to be included in the system in the second and third building blocks.

For example, a *modus operandi* retrieval module is already under development. *Modus operandi* is one of the more promising types of leads for the solution of crime but, until now, no manual system has been developed that comes close to exploiting its full potential. Computer processing promises a major breakthrough. The New York State Identification and Intelligence System project is engaging in system analysis aimed at achieving a significant improvement in this area through the use of computerization.

A detailed criminal history of each subject would be included in a central file, bringing together selected information now stored in the separate files of the different agencies that contact criminals during the processes of criminal justice. This capability would give every type of agency access to a greater quantity of information than they now have in their own files, thus facilitating the information sharing between agencies that already is necessary. This file, continually enriched from all sources, would cover the history of the criminal's contact with the police, district attorney, court, office of probation, correctional institution, and parole officer, and it would supplement information in the arrest-disposition capability.

Another capability offering great promise would be a fast-response warrant-and-wanted file of those persons who are wanted within the State, including missing persons and persons for whom warrants have been issued. In addition to the police agencies and the district attorneys, courts and parole and probation agencies would be assisted by this capability.

Another module being given high priority for early inclusion in subsequent building blocks is a file of intelligence information pertaining to organizations, people within organizations, and to selected criminals. Organized crime and group-criminal activity, as well as information on subversive activities, would be a special focus of such a module.

Other very useful capabilities will be added. Such areas as stolen motor vehicles, stolen property, property marks, forgeries of automobile registration and stocks and individual social histories as well as scientific data, will be considered. Most of these areas will be planned in the early stages of the system. Considerable system analysis will be required to develop approaches for these modules that may provide important advances over current practices.

#### THE DEVELOPMENT PROCESS

The production of a complicated electronic data processing system involves many steps: (1) System analysis, (2) system design, (3) production, (4) data conversion, (5) test and installation, and (6) implementation.

##### System analysis

This initial step will not only provide a detailed definition of the specific needs of the various local and State agencies, but will also spell out the improvement in service that can be provided for the agencies by blending electronic information processing technology with improved manual procedures. Next, detailed requirements for the system will be specified and translated into a suitable plan for development, and finally, a plan for production and implementation of the electronic data processing system will be established. Development of any new techniques that may be required will take place in this phase, also. The following distinct tasks are involved in this total process:

- Analysis of existing system,
- Development of new techniques,
- General requirements for the proposed system,
- Preparation of a plan for the overall system development efforts, and
- Production of system operational requirements.

##### System design

The general system operational requirements, phrased in user-oriented language, will be subjected to detailed analyses oriented toward computer programming to produce a set of operational design requirements in technical language, specifying those functions and activities to be performed by men and those to be performed by machine. Similarly, organizational requirements to effectively

implement personnel training requirements will be developed. Finally, all of the aforementioned products will be integrated into the blueprints for the system—system design specifications. This document will specify and design the input and output formats and the flow of information both internal and external to the computer. From these system design specifications, detailed program design specifications, defining specific computer program functions, will be prepared.

##### Production

In this step, the actual information-processing components—computer programs, operating procedures, and implementation handbooks—will be generated. All these components will be subjected to rigorous quality-control testing and debugging (detection and correction of errors). Upon completion of these activities, the nonequipment portions of the system will be available, and actual system installation can begin.

##### Data conversion

At the same time that the system design and production are taking place, the data conversion (from manual to machine form) will be planned and implemented. In building block No. 1, this will involve the transcription of existing material on criminal history to a computer-readable format for the arrest and disposition module. A selection of fraudulent checks that are currently on file in agencies throughout the State will be made in order to uncover those most likely to be of help in the solution of subsequent crimes. These checks will then be coded according to the technique developed for the fraudulent check module and converted into a form acceptable to the computer. For the latent fingerprint module, a selection of prints will be made of major criminals involved in those types of crime where latent fingerprints are likely to be found. These selected prints will be classified according to the detailed single-print classification system chosen for the module and this, too, will be put in a computer-readable form. It is likely that the personal appearance module will require the establishment of new recording techniques. New forms and techniques will have to be developed and tested, and personnel trained to use them. In order to have a good backlog of personal appearance descriptions when the system becomes operational, all persons arrested for fingerprintable crimes in the State after the technique is available, should be classified according to the new forms.

##### Test and installation

This step will include the assembling and integrating of system components, as well as extensive system testing. Furthermore, men, machines, procedures, and organization must receive a final integration through experience with the system.

##### Implementation

This phase is concerned with the orientation of personnel designated by New York State in the use of the electronic data processing system; it also includes preparation of necessary instructional materials.

The New York State identification and intelligence system—

- Is voluntary on the part of all agencies;
- Is not a competing agency; it will have no operational powers such as arrest, prosecution, or confinement;
- Will not interfere with the necessary autonomy of operation of these agencies;
- Will minimize the duplication of effort among agencies;
- Will realize substantial advances in the protection of civil liberties;
- Will provide unity among the agencies administering criminal justice where unity is appropriate;
- Will allow new levels of service to the participating agencies;
- Will make facts commonly used by all six functional areas in the administration of criminal justice readily accessible, as well as certain types of facts especially valuable to particular functions;
- Will make available, on a "right-to-know" basis, to each individual agency the equivalent of all relevant files in the State;
- Will contain devices allowing the transmission of photos, prints, handwriting samples, and other material;

- Will have a constantly enriched and updated central data base;
- Will contain in its data base only information used in the administration of criminal justice;
- Will better serve the public interest and better use public funds by the mechanisms it provides for information sharing among the agencies of criminal justice;
- Will be operated by present personnel on premises of user agencies;
- Will be closed to all but a defined category of public officials and agencies; and
- Will maintain security; highly sensitive data will be tagged in the system to insure that its distribution is under the exclusive control of the contributing agency.

## IN BRIEF

The New York State identification and intelligence system project was conceived to improve the processes of criminal justice within the State. However, its potential is exciting interest and inquiries from sister States and foreign countries confronted with the same challenges to their existing systems for the administration of criminal justice. It represents an historic event in that, so far as can be determined, it is the first time that every type of agency engaged in the administration of criminal justice has been brought together to think out its common information needs and problems.

Out of this meeting of minds has arisen a pioneering effort to systematically investigate the information needs of all the agencies and to meet those needs by the application of modern advances in electronic information processing technology.

New York State and System Development Corp. are jointly embarked in this effort. Through a mixed-team approach and a tested production methodology, invaluable tools will be developed to aid these agencies in carrying out their responsibilities. When completed, the system will provide faster services to the participating agencies by means of accurate and timely information transmission. The system will permit State and local agencies to substantially improve their information sharing and expand their operational capability.

## STATE OF NEW YORK

Print. 5333

Intro. 5104

## IN ASSEMBLY

March 18, 1965

Introduced by Mr. DOWD—read once and referred to the Committee on Ways and Means

## AN ACT

To amend the executive law, in relation to the creation of the New York state identification and intelligence system in the executive department and prescribing its powers and duties

*The People of the State of New York, represented in Senate and Assembly, do enact as follows:*

- 1 Section 1. The executive law is hereby amended by inserting
- 2 therein a new article, to be article twenty-one, to read as follows:
- 3 *ARTICLE 21*
- 4 *STATE IDENTIFICATION AND INTELLIGENCE SYSTEM*
- 5 *Section 600. Legislative findings and objectives.*
- 6 *601. Definitions.*
- 7 *602. Creation of the system; appointment of director.*
- 8 *603. Functions, powers and duties of the system.*
- 9 *604. Functions, powers and duties of the director.*
- 10 *605. Assistance of other agencies.*

EXPLANATION — Matter in *italics* is new; matter in brackets [ ] is old law to be omitted.

## 2

1 606. Studies; surveys.

2 607. Reports.

3 608. Grants or gifts.

4 § 600. Legislative findings and objectives. The legislature hereby  
5 finds and declares that:

6 (1) The sound administration of criminal justice importantly  
7 depends upon the effective collection, assimilation and retrieval  
8 of available information and its dissemination to appropriate  
9 agencies of government;

10 (2) It is in the public interest that, to the greatest extent pos-  
11 sible, government agencies concerned with the detection, apprehen-  
12 sion, prosecution, sentencing, confinement and rehabilitation of  
13 criminal offenders share among themselves available information  
14 relating to such offenders;

15 (3) At this time, relevant information is contained in many  
16 separate and widely dispersed file systems, manually maintained by  
17 government agencies throughout the state, and no adequate system  
18 now exists for coordinating either the files or the information they  
19 contain;

20 (4) There is a need to improve substantially the coordination  
21 of relevant information and to assure that it is disseminated accu-  
22 rately and swiftly, especially in aid of police officials, prosecutors,  
23 criminal courts, and probation, correction and parole officials;

24 (5) Through the use of electronic data processing and related  
25 procedures, a system should be established to provide a central data  
26 facility, by which relevant information can be coordinated and  
27 made readily available whenever and wherever required in the

## 3

1 investigation and prosecution of crime and the administration of  
2 criminal justice.

3 § 601. Definitions. As used in this article:

4 1. The term "system" means the New York state identification  
5 and intelligence system.

6 2. The term "director" means the director of the New York state  
7 identification and intelligence system.

8 3. The term "qualified agencies concerned with the administra-  
9 tion of criminal justice" means courts of record, probation depart-  
10 ments, sheriffs' offices, district attorneys' offices, state division of  
11 parole, New York city parole commission, state department of cor-  
12 rection, New York city department of correction, and police forces  
13 and departments having responsibility for enforcement of the  
14 general criminal laws of the state.

15 § 602. Creation of the system; appointment of director. There  
16 is hereby created within the executive department a New York  
17 state identification and intelligence system. The head of such sys-  
18 tem shall be a director, who shall be appointed by the governor, by  
19 and with the advice and consent of the senate, and shall hold office  
20 during the pleasure of the governor. He shall receive an annual  
21 salary to be fixed by the governor within the amount available  
22 therefor by appropriation. He shall also be entitled to receive  
23 reimbursement for expenses actually and necessarily incurred by  
24 him in the performance of his duties. The director may appoint  
25 such officers, employees, agents, consultants and special committees  
26 as he may deem necessary, prescribe their duties, fix their compen-  
27 sation and provide for reimbursement of their expenses within  
28 the amounts available therefor by appropriation.

4

1 § 603. *Functions, powers and duties of the system. The system*  
 2 *by and through the director or his duly authorized officer or*  
 3 *employee shall have the following functions, powers and duties:*  
 4 (1) *to establish, through electronic data processing and related*  
 5 *procedures, a central data facility with a communication network*  
 6 *serving qualified agencies concerned with the administration of*  
 7 *criminal justice located anywhere in the state, so that they may,*  
 8 *upon such terms and conditions as the director and the appropriate*  
 9 *officials of such qualified agencies shall agree, contribute informa-*  
 10 *tion and have access to information contained in the central data*  
 11 *facility, which shall include but not be limited to such information*  
 12 *as criminal record, personal appearance data, organized crime*  
 13 *intelligence, fingerprints, photographs, handwriting samples and*  
 14 *other related data;*  
 15 (2) *to receive, process and file fingerprints, photographs and*  
 16 *other descriptive data for the purpose of establishing identity and*  
 17 *previous criminal record;*  
 18 (3) *to adopt such measures to assure the security of the system*  
 19 *as the director deems appropriate;*  
 20 (4) *to engage in research and make studies and analyses of the*  
 21 *problems of identification and intelligence and to make the results*  
 22 *thereof available for the benefit of municipalities and state agencies*  
 23 *as the director may deem appropriate;*  
 24 (5) *to do all things necessary or convenient to carry out the*  
 25 *functions, powers and duties set forth in this section.*  
 26 § 604. *Functions, powers and duties of the director. The director*  
 27 *shall be the chief executive and administrative officer of the sys-*

5

1 *tem and shall, subject to rules and regulations approved by the*  
 2 *governor, direct the work of the system.*  
 3 § 605. *Assistance of other agencies. To effectuate the purposes*  
 4 *of this article, the director may request and receive from any*  
 5 *department, division, board, bureau, commission or other agency*  
 6 *of the state or any political subdivision thereof or any public*  
 7 *authority such assistance, information and data as will enable the*  
 8 *office properly to carry out its powers and duties hereunder.*  
 9 § 606. *Studies; surveys. In the accomplishment of the purposes*  
 10 *of this article, the director may undertake research and studies*  
 11 *through the personnel of the system or in cooperation with any*  
 12 *public or private agencies, including educational, civic and research*  
 13 *organizations, colleges, universities, institutes or foundations.*  
 14 § 607. *Reports. The system shall from time to time report to the*  
 15 *governor, and shall make an annual report to the governor and the*  
 16 *legislature, not later than May first, concerning the work of the*  
 17 *system in the preceding calendar year.*  
 18 § 608. *Grants or gifts. The director, with the approval of the*  
 19 *governor, may accept as agent of the state any grant, including*  
 20 *federal grants, or any gift for any of the purposes of this article.*  
 21 *Any moneys so received may be expended by the system to effectuate*  
 22 *any purpose of this article, subject to the same limitations as to*  
 23 *approval of expenditures and audit as are prescribed for state*  
 24 *moneys appropriated for the purposes of this article.*  
 25 § 2. *This act shall take effect April first, nineteen hundred*  
 26 *sixty-five.*

Mr. GALLAGHER. We would hope that when we have some further hearings on this that your own experiences in this field will have broadened to an extent that will be again even more helpful to us in the future as we consider this problem. I want to compliment you for an excellent presentation. I want to compliment you, too, for your civic spirit in offering to give us the benefit of your advice and of the program that you have instituted in New York State. I think it is extremely useful and will be helpful to us in our future deliberations.

I must run and if you do not mind, our counsel and administrative assistant might have a question or two. Subject to that, the committee would then adjourn.

Mr. GALLATI. I would like to express my sincere thanks to you. You have been most courteous and I appreciate the opportunity of being of any assistance I might have been.

Mr. GALLAGHER. You have been, and we certainly appreciate your appearance.

Mr. GALLATI. Thank you.

Mr. GALLAGHER. For the record, this will be an informal exchange, if you don't mind, with our staff, because theoretically I would now adjourn.

Mr. CORNISH. Is the State Civil Service Commission going to be one of your users?

Mr. GALLATI. The answer to your question, sir, is yes, the State Civil Service Commission would be a user of the NYSIIS file. I might add they are looking forward to this. They anticipate a considerable degree of relief from much of the type of investigation that they now conduct. This type of investigation that they conduct will be assisted by NYSIIS so that a great deal of the annoyance to the civil service applicant will be avoided through getting ready information from NYSIIS.

Mr. CORNISH. Will all applicants for jobs be checked against the system or just certain occupations?

Mr. GALLATI. This would be within the provisions of civil service and appointing officers. We would be recipients of the request. Since it comes from an appropriate source, it would be in accordance with their regulations, the details of which I am not familiar with.

Mr. CORNISH. I think I told you on the telephone the experience they had in the State of Ohio on this where they were checking applicants who wanted jobs as attendants in State mental hospitals in Ohio, and they had a criminal identification center—it was not computerized—at one of the State prison farms. They utilized convict labor as clerks in that office there. Sometimes a request would come in for an identification on one of the old buddies of somebody who was working in the office and they would very conveniently file that one away or put down a negative answer as to his previous criminal record. I assume you are not going to get into any problems like that.

Mr. GALLATI. I can assure you, Mr. Cornish, that we will check out most carefully all those employees that we have working on the system. As a matter of fact, we are now doing so and have done so in the past. This was so even before NYSIIS assumed responsibility for the division of identification. They are very carefully checked for this very reason that you mentioned.

Mr. CORNISH. Individuals picked up as suspicious persons, would they be listed in these records?

Mr. GALLATI. No; definitely not.

Mr. CORNISH. Does New York State law provide in any way that a convicted criminal can become later on, after his supposed rehabilitation, a State employee under any conditions?

Mr. GALLATI. Yes; our provision is in the State law for executive pardons, clemency, and also certificates of relief from disabilities. It is the policy of the Governor to encourage this type of rehabilitative therapy which is involved in employing people who do have backgrounds which have been relegated to a lesser position because of their subsequent good conduct.

Mr. CORNISH. There is some official recognition on the part of the State that a person can rehabilitate himself even though he may be a criminal.

Mr. GALLATI. Yes; definitely. This is not to say on the obverse side of the ledger that the sensitive job should be filled by persons with this type of background.

Mr. CORNISH. You may be interested to know that even in the case of people who have suffered mental illnesses, for example, the Federal Civil Service Commission has been interested in seeing what types of useful Federal employment that some of these people can perform within their limited means.

Mr. GALLATI. It certainly seems like a worthy endeavor.

Mr. CORNISH. In the context of these computer systems these questions have some relevance, as you can well imagine, because the computer system can make a very handy recitation of a person's former record, whether it be a criminal violation or whether it be a mental disorder.

Now, on another subject. You may have covered this earlier when I was out of the room, but did you describe at all your special relationship to the New York City Bar Association study group on privacy?

Mr. GALLATI. I did not go into that directly.

Mr. CORNISH. I think it would be useful if you could give us a brief description of that.

Mr. GALLATI. I would like to, if I may, point out that we have had the privilege of discussing the system with the special Committee on Science and Law of the Association of the Bar of the City of New York, and they have viewed our explanatory movie and listened to Mr. Lumbard and myself describe the system. They responded in a manner which was most helpful to us.

We have had subsequent conferences with members of that committee and also with Prof. Alan Westin, who is engaged as a staff consultant and writer for the committee, and we are attempting to build into our whole context those types of things which they felt were relevant to the proper development of this type of a system. They have been most helpful.

I might add that we have also discussed the system with a number of other people and agencies and institutions who are likewise quite helpful in alerting us to the types of problems we might anticipate. I would like to say this committee will also help to do this very thing for us.

Mr. CORNISH. Any of the suggestions they made, were they worked into this system at all?

Mr. GALLATI. Yes; they are all being taken into consideration and as we develop through the hierarchy, or the rank order, if you will,



of security as we get into more and more sensitive modules, each one of these will be taken into consideration at the appropriate time and given very serious consideration, and we will, of course, make every effort to build these suggestions into the system.

In this formative stage we are now in it is very important for us to get this type of influence because I think we can then overcome this tremendous danger of setting in concrete something which can later not be readily corrected.

Mr. CORNISH. So this liaison which you have with them is a helpful thing?

Mr. GALLATI. It is most helpful.

Mr. CORNISH. The reason I ask that question is, the chairman has proposed a similar liaison between the Bureau of the Budget in the establishment of a national data center, and also persons who would have expertise in questions of constitutional law and the special questions involving invasion of privacy.

I am glad to hear that that association was fruitful.

I do not have any more questions. Perhaps Mr. Romney and Mr. Forsyth do.

Mr. ROMNEY. I have one, Mr. Gallati.

How many employees do you have now working in your system, how many contractor employees, and how many do you expect to have when the system is fully operational?

Mr. GALLATI. We now have on board just a few persons over 300. These are people who are involved in the present manual, to be later computerized, operations of the old division of identification of the department of correction, which we assumed as of April 1 of this year. It includes also more than a hundred people who are involved in a massive data conversion effort.

We are taking the criminal records which are now maintained in manual form and putting these into computerized forms. The effort is tremendous because it is one of the largest data conversion efforts that have ever been undertaken in the State of New York, and perhaps the largest of this type that has ever been done in the world. These people are doing this type of work.

Then we have built up a group of professionals who are on board, State employees systems analysis people. We have with us a topnotch criminalist and we have a number of people who are experts in the field of system design and in terms of analysis of future modules.

And we have beside me here today my executive assistant, also a former member of the New York City Police Department, who is about to receive his Ph. D. in public administration, and has been very helpful in terms of this very area which we are considering today, the area of the right of privacy and civil rights and civil liberties generally.

Mr. ROMNEY. I think you should identify him by name.

Mr. GALLATI. This is Mr. Edward de Franco, also a patrolman of the city of New York on leave like myself.

I think also a point should be made in terms of the reception of the system by the people of the State of New York, and by the officials both in the legislature and in the executive chambers, that we have had a response from the grassroots, and particularly from the people who are in the professional fields with which we will have to deal, which

has been most heartening and perhaps beyond our wildest dreams when we began the system. Their support has been expressed by their representatives in the assembly and the senate who have voted rather substantial budgets for this purpose, and when the bill creating NYSIS, which will be part of the record, was before the assembly and the senate, the vote on that was 57 to 0 in the senate and 127 to 1 in the assembly—and this is in a State which I need not remind you is very conscious of matters relating to the types of things which we are addressing ourselves to here today.

In addition to the State employees which I mentioned, we also have a cadre of seven people from the System Development Corp., who have been with us from the very beginning of the development of the system and who have come along with us and are familiar with all aspects of it.

Among these we have a gentleman by the name of Eric Witt, who is full time, as I said before, assigned to these aspects of development of the system.

The other contractors we have are in the area of data conversion. We have a consultant in this area. Touche, Ross, Bailey & Smart is the consultant we are using in this area. We are using Computer Usage for some of our programing. We are using Jansky & Bailey for the evaluation of our facsimile network and communications matters generally.

We have also employed the technical assistance branch of the Hughes Aircraft Corp. for studies in the intelligence area.

I believe that is the complete list of those who are our corporate consultants, at least as far as I can recall them at the moment.

Mr. ROMNEY. When your system is operational you will have changes of course in your personnel requirement. But do you have an estimate of the personnel that you will be using then?

Mr. GALLATI. We have estimates, but these are of course not truly scientific at this time because it is difficult to gage just what the relationship will be between the savings in terms of manpower by utilizing computerized techniques and the additional pressures we will have to continue to expand on our modular development.

I anticipate, for example, as we develop this single fingerprint capability, instead of receiving as we now do 3 latent prints; that is, 3 scene-of-the-crime prints per week from the entire State of New York, this may be magnified in the order of 50 a week or perhaps a greater number.

Budgetwise, which I mentioned in terms of the support we have had from Governor Rockefeller and the executive chamber and from the "grassroots," expressed in the votes of the legislature, we had in the beginning, in 1963, for the original feasibility study, an appropriation of \$50,000. In 1964, we received an appropriation of \$500,000; and last year we had an appropriation of \$1,250,000. We are now entering the new budget period with a budget of \$3,777,000, for which we are most grateful to the people of the State of New York and to the government of the State of New York.

I think with this type of support we can develop a rational and good system which will serve the interests of the community.

We have, in the area of the support which we have received, another aspect which is less direct but one which we appreciate very

much also. That is in the area of comments we have heard from knowledgeable people in other parts of the world and also from that most sincere form of flattery, which is imitation—a number of States are considering the type of system which we have for their States, and also in terms of scientific reports that we have received.

We recently received such a report from the home office of the United Kingdom in which they refer very flatteringly to the method in which NYSIIS was developed as the most thorough development of computerized systems for agencies of criminal justice they can conceive, and one which they commend for the consideration of the Queen.

Mr. ROMNEY. Is Ohio one of those States?

Mr. GALLATI. I am reminded we are not sure just yet how far or in what direction these systems will go. However, we are made aware of the fact that Ohio is interested in this type of system and may well take the same course we have.

We would like to make, and I think it is important to note, the distinction between some systems which exist of this type in various parts of the United States and our system. Many of these systems that exist in other States are police systems, they are under the jurisdiction of the State police or of the county police department, and they are dedicated to law enforcement, as opposed to the entire spectrum of criminal justice and as opposed to our concept of a totally disinterested agency which has no arrest, investigative, or custodial responsibility.

Mr. FORSYTH. I would like to ask a question on that point. I know you have set up this central system on that concept. Did you have security reasons as substantial consideration when this was done?

Mr. GALLATI. This was one of the reasons.

I think to really answer your question properly, I would have to refer also to the basic common acceptance of the system as a new concept in State government. You do have in New York State—

Mr. FORSYTH. Excuse me. I just wondered if it was a substantial consideration of the setting up of the system independently, for security reasons alone.

Mr. GALLATI. This was a consideration, but I would not say it was the controlling situation. I think perhaps the most substantial reason might be the fact that we did have in New York State, as I assume may be the case in other States, a certain amount of lack of confidence between and among the various agencies of criminal justice.

Mr. FORSYTH. What I am concerned with primarily is the security aspects. Probably it was not really the controlling aspect, a substantial aspect, but you did consider it?

Mr. GALLATI. Yes.

Mr. FORSYTH. Do you think centralizing it increases or decreases security problems it had?

Mr. GALLATI. I think it has decreased the problems of security by centralizing in this fashion.

Mr. FORSYTH. Can you tell me in a few short words why you think it decreases them?

Mr. GALLATI. In the first place, the fact that it was an independent agency, which therefore had no axes to grind, no utilization for the information which was introduced to it other than to properly handle it in the fashion of—perhaps one might use the example of a trustee

who is not utilizing this information for his own purposes but is utilizing it for the people to whom he is responsible.

Secondly, in reference to the amount of security that can be afforded in a central location such as this, the fact that information is filtering in which may be much more sensitive at a local location and becomes less sensitive by removal from the local area, will have a bearing upon it. But, more particularly, I think as you get into an area of larger responsibility, you necessarily will apply more resources to the protection of this responsibility and since the State does have the capabilities and the resources to apply to it, I am sure that we can afford greater security all the way along the line. I keep thinking of Fort Knox as opposed to some drive-in bank. Perhaps this is a bad analogy but I think that we can provide a greater amount and quality of resources to secure the information than can a smaller community which might be limited in its resources.

Mr. FORSYTH. Thank you.

Mr. GALLATI. I might say, of course, NYSIIS is a service agency only. It exists for this purpose, to serve the people who are in the professions related to criminal justice.

Mr. CORNISH. Once again, on behalf of the staff, we want to thank you for staying and answering a few additional questions.

Mr. GALLATI. We are delighted to have the opportunity and appreciate your courtesy and consideration.

(Whereupon, at 1:30 p.m., the committee adjourned. Mr. Gallati later submitted a supplementary statement and additional documents which follow:)

#### SUPPLEMENTARY STATEMENT OF DR. ROBERT R. J. GALLATI

There is an increasing requirement for greater scientific capabilities and more extensive and sophisticated information sharing among criminal justice agencies within each State. There is also a great need existing in State and local agencies to receive information from the various Federal agencies of criminal justice. Indeed, there is believed to be a great need for Federal criminal justice agencies to share information with each other. Greater capabilities in this area do not necessarily imply greater risk for society; in fact, the opposite may well be true.

The perils we anticipate from large collections of data exist today in our inefficient manual files. It is to the credit of this special subcommittee that these risks are being exposed for rational evaluation. It is also important to recognize that the potentials of computerization have stimulated the concern which exists here today—a concern which is long overdue.

The computer has compelled society to consider the value of security systems applied to information entrusted to data centers, whether they be manually operated, or otherwise. In this moment of truth, the doctrine of the right of privacy has come to share a new place of honor among other protected rights. We need to protect private personality as zealously as we protect private property, for as we protect the right of privacy we protect the right to share and communicate. Surely, privacy protected implies parameters of when, where, with whom, what, how and why information should be shared or withheld. As with all rights, a paramount public interest, accepted by the community and explicitly recognized, should equate, in productive equilibrium, the claim of privacy and the need to share information.

It is within this context of security and a climate of concern for the protection of individual rights and liberties that the New York State Identification and Intelligence System (NYSIIS) is being developed. NYSIIS is a voluntary system of information sharing to serve the agencies of criminal justice in New York State. NYSIIS has no arrest or investigative powers or responsibilities. It is an independent agency created solely to serve criminal justice and to store and retrieve files and forms relevant to the criminal justice process. Thus, information such as contained in census, tax, election, social security, unemployment insurance and similar files would not be collected. We would not expect

that anybody would submit the names of confidential informants, wire tap information, grand jury minutes, etc. It is not planned to include family court information in the NYSIS data base.

REPORT OF THE SPECIAL COMMITTEE ON PROBLEMS OF LAW ENFORCEMENT OF THE ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK ON PROPOSED LEGISLATION TO AMEND THE EXECUTIVE LAW IN RELATION TO THE CREATION OF THE NEW YORK STATE IDENTIFICATION AND INTELLIGENCE SYSTEM IN THE EXECUTIVE DEPARTMENT, AND PRESCRIBING ITS POWERS AND DUTIES

(S. Int. 3547, Pr. 3960, 5041, Mr. Mangano; A. Int. 5104, Pr. 5333, 6519, Mr. Dowd)

By report dated January 2, 1964, this committee recorded its support of the basic proposition that efficiency and effectiveness of law enforcement will be enhanced by increased cooperation by the various agencies charged with law enforcement responsibilities, and its view that the pooling of information with respect to all matters of concern to such agencies is an activity where such cooperation is particularly appropriate. This report strongly endorsed the feasibility report and recommendations for a New York State Identification and Intelligence System which was designed to utilize electronic data processing equipment for the gathering, recording, and dissemination of criminal intelligence to law enforcement agencies throughout the State.

The proposed legislation (S. Int. 3547, Pr. 3960, 5041; A. Int. 5104, Pr. 5333, 6519) is designed to effectuate the program recommended in the feasibility report. It authorizes the appointment of a director of a proposed New York State Identification and Intelligence System and provides the director with authority to establish a central data facility with a communication network through which law enforcement agencies, as well as other agencies concerned with the administration of criminal justice, can coordinate the collection and dissemination of criminal justice.

This committee believes that the proposed legislation, by facilitating the collection and exchange of criminal intelligence, constitutes a major step forward in improving cooperation among law enforcement agencies, and in thereby improving the effectiveness of law enforcement activity throughout the State. Accordingly, we approve the proposed bills and recommend their prompt adoption.

May 19, 1965.

SPECIAL COMMITTEE ON LAW ENFORCEMENT, THE ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK

Robert B. Fiske, Jr., Chairman.  
Arnold Bauman  
Edward Q. Carr, Jr.  
Arthur H. Christy  
Raymond L. Falls, Jr.  
Fred N. Fishman  
Victor S. Friedman  
Grenville Garside  
Edwin L. Gasperini  
George I. Gordon

David Klingsberg  
Stephen P. Kennedy  
Whitman Knapp  
Newman Levy  
Jerome J. Londin  
Charles H. Miller  
Robert P. Patterson, Jr.  
Leonard B. Sand  
Leon Silverman  
Paul Windels, Jr.

REPORT OF THE SPECIAL COMMITTEE ON PROBLEMS OF THE ADMINISTRATION OF LAW ENFORCEMENT OF THE ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK ON THE FEASIBILITY REPORT AND RECOMMENDATIONS FOR A NEW YORK STATE IDENTIFICATION AND INTELLIGENCE SYSTEM

The Committee on Problems of the Administration of Law Enforcement of the Association of the Bar of the City of New York is of the opinion that the efficiency and effectiveness of law enforcement will be enhanced by increased cooperation among the various agencies charged with law enforcement responsibilities. This committee believes that the pooling of information with respect to all matters of concern to such agencies is an activity where such cooperation is particularly appropriate. Such pooling will eliminate wasteful duplication and at the same time increase the total fund of information available to each

participating agency. The techniques to be employed by the State in the gathering, recording, and dissemination of such information should utilize recent technological advances in data processing. The feasibility report and recommendations for a New York State identification and intelligence system which has been submitted to the New York State executive chamber by the System Development Corp. concludes that a system for the rapid collection, collation, and dissemination of such material from a central electronic data processing system is completely feasible from a technological standpoint and recommends that New York State proceed with the study and system design necessary before such a system may be produced and placed in operation. Although the members of the committee are not competent to pass on the technological aspects of the proposed system, the committee is of the strong opinion that the entire study is well warranted and should proceed with dispatch. We assume that in the course of this study consideration will be given to technological problems and to development of safeguards to preclude any misuse of the system.

January 2, 1964.

Robert B. Fiske, Jr., chairman; Arnold Bauman; Edward Q. Carr, Jr.; Arthur H. Christy; Raymond L. Falls, Jr.; Fred N. Fishman; Victor S. Friedman; Grenville Garside; Edwin L. Gasperini; George I. Gordon; David Klingsberg; Stephen P. Kennedy; Whitman Knapp; Newman Levy; Jerome J. Londin; Charles H. Miller; Robert P. Patterson, Jr.; Leonard B. Sand; Leon Silverman; Paul Windels, Jr.

THE ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK

COMMITTEE ON CRIMINAL COURTS

(1965 Legislative Bulletin No. 12)

(A. Int. 2625, Pr. 2631, Mr. Curran; A. Int. 5104, Pr. 5333, Mr. Dowd; S. Int. 1481, Pr. 1495, Mr. Bush; S. Int. 3647, Pr. 3960, Mr. Mangano)

Identification and intelligence system, creation of within executive department (add new Exec. L. Art. 21).

APPROVED

These bills would create a statewide identification and intelligence system within the executive department.

The bills provide for a director, to be appointed by the Governor with the advice and consent of the senate, and to hold office during the pleasure of the Governor. They would establish a central electronic data processing facility which will receive and disseminate information important in the administration of criminal justice—identification data, criminal records, organized crime intelligence—pursuant to agreements reached with courts of record and the various qualified law enforcement and correctional agencies throughout the State.

The technical feasibility of the proposed system and its value to the relevant agencies were explored during 1963 in a study authorized by the Governor, with the concurrence of the legislature, and were favorably reported upon in November 1963 with many of the agencies being represented in an advisory capacity particularly those in New York City.

Since it appears to be feasible and valuable, we recommend that the system be put into effect as swiftly as possible.

For the reasons stated, these bills are approved.

COMMITTEE ON CRIMINAL COURTS, LAW AND PROCEDURE, THE ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK

Richard A. Green, chairman  
Harvey P. Dale  
Arthur L. Liman  
Irving Mendelsohn  
Courtlandt Nicoll  
Burton B. Roberts  
Francis L. Valente  
Thomas D. Edwards  
William Scott Ellis  
Eleanor J. Piel (Mrs.)

Leonard Reisman  
Irving Younger  
Edward Brodsky  
Malachy T. Mahon  
Leon Polsky  
Harold Reynolds  
Edwin Silberling  
H. Richard Uviller  
Patrick Wall

(The following material was received for the record:)

STATEMENT BY JOHN DE J. PEMBERTON, JR., EXECUTIVE DIRECTOR, AMERICAN CIVIL LIBERTIES UNION, NEW YORK, N.Y.

The American Civil Liberties Union is grateful for the opportunity to present its views regarding proposals for a Federal Data Center or centers which the Special Subcommittee on Invasion of Privacy is now considering. We applaud the Subcommittee's concern over values threatened by these proposals and the observation by Congressman Gallagher that they raise "serious questions about individual rights."

The implications of such proposals shock the sensibilities of thinking Americans. In our modern age, with all of its intrusive impact on the individual, traditional concepts of a man's right to privacy are already being increasingly undermined. These proposals would alarmingly accelerate this trend.

The privacy of which we speak has been defined in these terms:

"The essence of privacy is \* \* \* the freedom of the individual to choose for himself the time and circumstances under which, and most importantly, the extent to which, his attitudes, beliefs, behavior and opinions are to be shared with or withheld from others." Rueblhausen and Brim, "Privacy and Behavioral Research," 65 Colum. L. REV. 1184, 1189 (1965).

It is our concern that this privacy, which is an essential source of civil liberties and therefore lies at the heart of our free society, will seriously be jeopardized by proposals presently being made for the collecting and centralizing of all data possessed by the Government regarding each citizen.

We concentrate our comments on two proposals, the FBI National Crime Information Center and the National Data Center.

Under the first proposal, a network of computers would store all information on criminal conduct. Certain valid law enforcement purposes will be served by the creation of such a data center. Police work and crime detection can be more efficiently pursued if information concerning major crimes is readily and quickly available to law enforcement officials. In addition, such a center can serve as a source of vital statistical research on crime and police practices in the United States.

However, two dangers to civil liberties are inherent in the existence of such an information center. The first of these, the widespread use of incomplete and unexplained arrest records, has long concerned the American Civil Liberties Union. We have been deeply troubled by the adverse consequences to an individual flowing from the recording of an arrest not followed by indictment or conviction, as well as from the making of a record of certain arrests, and even convictions, where the true nature of the conduct leading to arrest (such as peaceful participation in civil rights or peace marches) is not disclosed. In our correspondence over the past few years with the FBI about the arrest record problem it has been clearly established that too frequently local law enforcement officials report arrests to the FBI but fail to report later disposition of the case. Countless persons against whom charges have been dropped or who have been acquitted must still suffer the harsh consequences of a wrongful taint of criminality when seeking employment or other privileges. These problems are even more grievous in the all-too-common case today of those arrested for the valid exercise of constitutionally protected rights. No reliable procedure exists for differentiating such arrests in present FBI records from arrests made for the normal incidents of criminal conduct.

The union has frequently suggested methods to eliminate employment discrimination based upon the bare record of an arrest, urged better reporting of arrests and ultimate disposition in each case, and pressed for avenues of legal redress for improper use of arrest records. Yet the problem remains, and will be accentuated by the creation of a central pool of information. Such a pool will serve only to multiply the deprivation of the civil liberties of those who are wrongly arrested or arrested and even convicted for merely exercising their rights. Inaccurate and prejudicial data will be made available to a greater number of police officials and through them to still greater numbers of unauthorized persons.

Our second concern regarding the proposed FBI Crime Information Center is that it will be the repository not just for crime information, which is a valid function, but for other types of information not at all relevant to the prevention and detection of crime. It is said that other Federal investigative agencies will

be invited to feed whatever information they choose into the huge reservoir that the national network of computers will store and retrieve. Data concerning a person's political beliefs and associations, gathered by various Federal security agencies, thus will become part of the crime data bank. The implications are obvious: every local police official will be able to learn with facility not only whether a suspect has a criminal record, a proper disclosure, but also whether he has at all deviated from his community's political or social norms, a highly improper disclosure which threatens the enjoyment of first amendment protections. We know from the history of our own McCarthy era how such information can be improperly used when placed in the wrong hands, to hound people out of their jobs and subject them to other reprisals for their dissenting or unpopular opinions.

We therefore urge that, should such an FBI Center be created, strong safeguards be instituted to insure that only proper or relevant information be stored and that it be used only for proper crime prevention or crime detection ends. Such safeguards should at a minimum—

1. Prohibit the receipt and storing of information other than that formally recorded in connection with the report of a crime and the commencement and disposition of a criminal proceeding.

2. Prohibit the disclosure of such information to any but an authorized officer of a law enforcement agency.

3. Effectively penalize the disclosure by the direct or ultimate recipients of information so obtained to anyone not authorized to receive it in the first instance, and

4. Prohibit the disclosure of any information concerning arrests made more than 60 days preceding disclosure.

5. Provide procedures for the subject individual to learn the whole of the record kept about him, and to compel correction of inaccuracies and prejudicial omissions and the striking of stale records of arrests that have not led to conviction.

Threatening as the proposed National Crime Information Center is, by far the more serious threat to civil liberties stems from proposals to establish a National Data Center whose ultimate design is to centralize in a single place all information possessed by any arm of the government on each individual. We agree with the statement of the chief personnel investigator for the Civil Service Commission that, "Whenever a bureaucracy amasses files about its citizens an inherent threat to liberty exists." In recent years, as our society has incredibly increased and proliferated information gathering about individuals, no person can embark on any of the activities which comprise the essence of modern life, from opening a charge account to seeking a change of jobs, without some investigator prying into his past and present life. In many instances these investigations serve a vital function. But far too often the quest for information expands out of proportion to the actual need. Worse, information relevant for one purpose may be disclosed in the course of a wholly different inquiry as to which it is both irrelevant and prejudicial.

There are two distinct threats to civil liberties posed by the creation of such a data bank. The first lies in the methods by which the information is gathered. Unfortunately, the great bulk of information about an individual is not gathered as the result of inquiries by skilled Government security investigators. Rather, it is often acquired by Government employees of poor judgment, by private agencies, credit unions, insurance companies, and businesses. Government agencies often farm out investigative work to private firms, and there is a considerable interchanging of data among Government and private sources. Once an unreliable bit of information makes its way into a file it forms an indelible mark on a person's record. The individual who is denied the chance for employment or some other opportunity on the basis of such information is given no chance to rebut or disprove it. Dossiers are compiled, the accuracy of which increasingly becomes more questionable. The computerization of such information in the data bank only compounds the basis abuse. Such procedures, taken as a whole, promote a society unresponsive to the necessities of human dignity and privacy, and in the particular case, deny the individual the elements of basic fairness.

The second threat posed by the existence of such a data bank is that information will be used in harmful ways irrelevant to the specific purpose in each case and not intended when the information was gathered. At present each of the numerous Federal agencies that gathered information on citizens maintain their own records for their own particular use. Under the proposed plan, all this information will be centralized. Ultimately, at the push of the proverbial button,



an official from any agency will have access to every bit of data ever accumulated on a person, including information completely unrelated to his concern. For example, information gathered from a passport application will be available to the official processing a request for an FHA loan. Medical and psychiatric histories of applicants for veterans benefits may become available for political misuse. In our view, others simply have no business seeing data gathered for such special purposes. That such access can be abused, and that improper determinations based on legally irrelevant information can be made, is a proved experience with Government. During the 1950's we learned the tragic lesson that the confidentiality of Government files is already too difficult to maintain and that there are unscrupulous persons who will utilize their access to file information for ulterior purposes.

There is danger that data gathered by the Government will find its way into the hands of private firms where it will be improperly used against an individual. Moreover, the reverse will also occur; thus, for example, a Government agency, itself unauthorized to administer a polygraph test to job applicants, will have available the results of such a test administered to the individual when he applied for employment with a private company. Prohibited results will be achieved in an indirect fashion. When any official determination is made on the basis of irrelevant information which that official has no right to consider, the end result is a deprivation of the individual's civil liberties.

In recent years the Supreme Court has erected constitutional barriers against the improper use of surreptitious surveillance devices (*Silverman v. United States*, 365 U.S. 505, 1961), has prohibited the attempted elicitation by Government agencies of information about an individual's associational ties (*Shelton v. Tucker*, 364 U.S. 479, 1960) and has taken the first step toward defining an area of protected privacy and sanctity into which the Government may not intrude (*Griswold v. Connecticut*, 381 U.S. 479, 1965). The proposed data bank threatens to vitiate these protections so vital to our free society. The content of conversations, monitored by eavesdropping devices, may well be codified and put on tape. The individual's associational ties will become part of his dossier available to the lowest personnel official. The private lives of our citizens, as "documented" by the basest form of gossip, will be fed into the machine. All this in the name of efficiency. This is Government surveillance which brings 1984 to our doorstep.

Confirming our apprehension over the alarming implications of this proposal is the absence, in descriptive statements we have so far seen, of emphasis upon any such vital safeguards as those that have been suggested to the subcommittee by Prof. Charles A. Reich—and the concomitant likelihood that such safeguards will be overlooked in its implementation. Such safeguards would include—

1. Prohibitions on the eliciting, storing, or retrieving by any officer of Government of certain types of information for any purpose.
2. Provision to each citizen of an effective right to inspect the contents of information collected about him and to compel corrections of its inaccuracies and prejudicial omissions, and
3. Limitations on the access to stored information to the original recipient of that information.

The seeming insensitivity of proponents of the National Data Center to the needs for such safeguards underscores the reasons for our opposition.

The principles which protect liberty and human dignity from the insistent demands of order and efficiency are the very stuff of which our democratic society is made. They demand that the collection, storing, and retrieving of essential information, such as that concerning crimes and criminal proceedings, be surrounded with safeguards of the kinds we have proposed. And they demand that all larger conceptions, looking to such essentially unlimited Federal surveillance as that of the proposed National Data Center, be rejected and discarded outright.

We commend the subcommittee for its vigilant attention to these hazards.

PAPER SUBMITTED BY DR. DONALD N. MICHAEL, PROFESSOR OF PSYCHOLOGY AND PROGRAM DIRECTOR IN THE CENTER FOR RESEARCH ON THE UTILIZATION OF SCIENTIFIC KNOWLEDGE, THE UNIVERSITY OF MICHIGAN ENTITLED "SPECULATIONS ON THE RELATION OF THE COMPUTER TO INDIVIDUAL FREEDOM AND THE RIGHT TO PRIVACY"

In this paper, we are concerned with the future—the next 20 years or so. To look even that far ahead may well be a futile exercise, for the rate of change of technology and society threatens to make footless fantasizing of any specula-

tions about the impact of selected factors. However, those who will have significant influence on the political and social processes of the next 20 years are alive today. Consequently, they share to some large degree the values prevailing now, and this is important if we want to explore the significance of the computer for our kinds of privacy and freedom. The years beyond the realm of a 20-year period, may find us dealing with a population a significant proportion of which holds values quite different from today's. Since we are not likely to know what those values might be, further speculations than we are about to undertake would hardly be worth the effort in the context of this Symposium.

One approach to the kind of speculations we shall pursue herein would be to review with great precision and perspicacity the history and ramifications of the concepts of privacy and personal freedom, and in this light, to look at the possible effects of computers on them. We will not follow this approach; space and the author's knowledge are too limited, and the concepts, whether they are refined philosophical, legal, ethical, or political formulations, undoubtedly will have their day when it is time to inhibit or facilitate the impact of the computer. Before then, the impact of computers on man will be reflected much more in the commonplace responses of our pluralistic society to these frequently misunderstood and misapplied concepts. In particular, it should be understood that the writer's grasp of these concepts is also of the "common" variety.

This paper should be read, then, as no more than a stimulus to further speculation and much hard work. It is a preliminary exercise, an attempt to delineate some circumstances where computers and the concepts of personal freedom and privacy may come together in the day-to-day environment of the next couple of decades to enhance or detract from the practice and preservation of freedom and privacy. We shall concentrate on the role of computers as the technological agents for these developments. The microphone, tape recorder, miniature camera, and questionnaire are other formidable technological agents; their uses are well documented in two recent books and we need not review the matter here.<sup>1</sup> We shall be concerned more with the implications for privacy and freedom implicit in the means and capacities of computers for processing and evaluating information, however collected. Our goal is to identify the interactions and the circumstances to look for if we wish to anticipate the impact of computers on freedom and privacy.

In order to grasp fully the potential impact of computers, we must be clear about our versatility. In their simplest forms they can sort punched cards and preform, at high speed, routine arithmetical and statistical calculations. In their more elaborate versions, computers can—

"be built to detect and correct errors in their own performance and to indicate to men which of their components are producing the error. They can make judgments on the basis of instructions programed into them. They can remember and search their memories for appropriate data, which either has been programed into them along with their instructions or has been acquired in the process of manipulating new data. Thus, they can learn on the basis of past experience with their environment. They can receive information in more codes and sensory modes than men can. They are beginning to perceive and to recognize \* \* \*. Much successful work has been done on computers that can program themselves. For example, they are beginning to operate the way man appears to when he is exploring ways of solving a novel problem. That is, they apply and then modify, as appropriate, previous experiences with and methods of solution for what appear to be related problems. Some of the machines show originality and unpredictability."<sup>2</sup>

Let us also recognize that the impact of computer technology will not be unilateral. Rather, it will be profoundly affected by attitudes held by significant portions of the public and their leaders—attitudes favorable, indifferent, or antagonistic to privacy and freedom. There are, of course, great social pressures already operating which run counter to the preservation of privacy. We shall not explore the sources of these pressures and anxieties; they are recognized as chronic states of mind and action for a large part of our population and its leadership. But they result in conformity and in the justification of exposure, and in order to conform or to assure that others meet certain standards of conformity, people need to know what other people are doing, especially in their less easily observable lives. Our mass media in particular stimulate and cater to this need, and they revel in the publicizing of personalities by stripping

<sup>1</sup> Packard, *The Naked Society* (1963) and Brenton, *The Privacy Invaders* (1964).

<sup>2</sup> Michael, *Cybernation: The Silent Conquest* 6-8 (1963).



away privacy, whether it be from the individual, his home, the classified senatorial hearing, or the diplomatic conference. There is every reason to believe that defining "reality" in terms of persons and in personal terms will continue, especially as the new, depersonalized reality becomes too complex to convey much meaning to the average citizen. Personalities are meaningful, and defining reality in terms of personalities will continue both to appeal to the conventional wisdom and experience of most people and to provide an attitudinal environment wherein it is more permissible for business and Government to probe persons, too.

Up to the present, "central city" concentration (in contrast to most suburban situations) population growth, and increasing physical mobility have given the individual some relative opportunity to lose himself, or to be anonymous, thereby preserving to some extent his privacy and freedom of action. As we shall see, much of whatever ecological advantage these sources of anonymity provide probably will disappear even if, as is unlikely, the flight to the suburbs ceases.

We must realize, too, that the ways in which the applications of computer technology affect other important aspects of our social environment inevitably will reinforce or overcome attitudes about freedom and privacy. In particular, the computer will have an increasingly significant influence on the design and conduct of public policies. The states of mind and conditions for action resulting from the implementation of these policies will affect the ease with which one can pursue freedom and privacy. An obvious inhibiting influence upon that pursuit would be produced by the siege style of command and control of society which Harold Lasswell calls the "garrison state."<sup>3</sup> A garrison state might well be the consequence of an ever more elaborate proliferation of national security policies, guided and embellished by the kinds of computer-based war games, weapons systems, and sophisticated strategies which have become fashionable in the last several years. On the other hand, a federally integrated attack on crime, fully using the ability of the computer to organize and interpret data about criminals and crimes, eventually would free many terrorized people from threats of death or disaster and open business opportunities now preempted by the freewheeling criminal. Thus, it would not be surprising if, in the future, people were willing to exchange some freedom and privacy in one area for other social gains or for personal conveniences. Nor would it be the first time they have done so.

With such background considerations in mind, let us speculate on particular circumstances in which the computer will confront what, in myth or actuality, we take to be present privileges of privacy or freedom.

#### PRIVACY

Consider that kind of privacy which exists by virtue of the ability to restrict access to information about oneself and one's related activities and records. By and large, the information thus restricted concerns the historical self: not only one's outward conduct, but also his inward evolution as a human being.

The availability of computers can alter seriously the degree to which one can restrict such access. Several factors which have determined degrees of privacy in the past are—

- (1) The ability of the privacy invader to bring together data which has been available, but which has been uncollected and uncollocated;
- (2) The ability of the privacy invader to record new data with the precision and variety required to gain new or deeper insight into the private person;
- (3) The ability of the invader to keep track of a particular person in a large and highly mobile population;
- (4) The ability of the invader to get access to already filed data about the private person; and
- (5) The ability of the invader to detect and interpret potentially self-revealing private information within the data to which he has access.

What is the interplay of these factors and what is their significance for privacy in the light of the computer's capabilities? Much of one's privacy remains undisturbed because no one has had the ability to pull together available information—or because no one has been sufficiently interested to go to the trouble of doing so. To understand the private implications in available data might first require both locating and integrating much widely dispersed information.

<sup>3</sup> Lasswell, "National Security and Individual Freedom," 47-49 (1950).

The meaning of the information may be unclear, and, therefore, still private. More information may be needed, and the quality of it may depend on updated surveillance of the person involved. Considering the size and mobility of our society, these problems have made privacy invasion very difficult, but, as we shall see, the computer makes it much more feasible.

Private information about a person may exist which is ethically or legally restricted to those who have a legitimate right to it. Such information, about a great portion of our population, exists in business, medical, government, and political files, and in the form of psychological tests, private and government job application histories, Federal and State income tax records, draft records, security and loyalty investigations, school records, bank records, credit histories, criminal records, and diaries. Each day more of these records are translated from paper to punchcards and magnetic tapes. In this way they are made more compact, accessible, sometimes more private, and, very importantly, more centralized, integrated, and articulated. The results are more complete records on each individual and a potential for more complete cross-correlations. The would-be invader who knows about these centralized or clustered inventories need not search for sources, and therefore he may be much more inclined to examine the records than if a major search for the sources of information were necessary.

As population and mobility increase, there will be other incentives to establish central data files, for these will make it easier for the consumer in new environments to establish who he is and, thereby, to acquire quickly those conveniences which follow from a reliable credit rating and an acceptable social character. At the same time, such central data files will make it easier for the entrepreneur or government official to insure his security, since he will know at all times with whom he is dealing. In consequence, we can expect a great deal of information about the social, personal, and economic characteristics of individuals to be supplied voluntarily—often eagerly—in order that, wherever they are, they may have access to the benefits of the economy and the government.

While this sort of information is accumulating, the behavioral scientist, in direct consequence of the capabilities of the larger computer, will be improving his ability to understand, predict, and affect the behavior of individuals and groups. For the computer provides two prerequisites for the development of effective social engineering. First, only the computer can process fast enough the enormous amounts of data needed to know what the existing states of social and economic affairs are. In the past, such information was cantay or non-existent, or what there was more or less out of date. This no longer need be; and it certainly will not continue to be, since coping with sheer social complexity will require that such information be abundantly available. In the second place, the computer will let the social scientist manipulate enough variables and enough circumstances in sufficiently complex ways to invent subtle models about the behavior of man and his institutions. Simulation of the behavior of individuals and institutions through the use of computers is well underway, and all signs are that it will be exceedingly productive. There is every reason to believe, then, that with the development of these sophisticated models, and with access to centralized data banks where many of the characteristics of each person, the institutions with which he is involved, and the environment in which he operates are recorded, it should be possible to develop a sophisticated understanding of the present behavior of individuals and to predict with some assurance various aspects of their future behavior as well as to interpret and deduce aspects of their past behavior. How detailed and valid the conclusions will in fact be remains to be seen, but it is very likely that average citizens, as well as those who will have a vested interest in using such predictions, probably will overestimate their precision. Even today, many people are willing, indeed sometimes want, to believe that the behavioral scientist can understand and manipulate their behavior. Whether the ability to predict behavior will be used to invade privacy and freedom will depend on more than technological capability, but certainly the capacity for invasion will increase as behavioral engineering progresses.

The ability of the user of private information to gain access to already collected data about the individual will depend on several factors. To understand the range of possibilities it should be recognized that, increasingly, data will be stored in memory banks shared by several users. A computer's memory banks can be so large that only a relatively few users can employ their entire capacity. Furthermore, the speed and capacity of computer processing is so high that it is much more efficient to have several users sharing time on

the machine. This means that while one user's finger is moving toward a control button on his computer control panel, another person is using the circuits he will be using a few seconds later. In the future, one computer user might accidentally gain access, through equipment malfunction, to another's information stored in shared memory banks. It is also possible that such information could be retrieved deliberately and clandestinely by querying the computer with someone else's retrieval codes or by otherwise tricking the computer and its memories. Shared computer time and shared memory banks are new techniques, and the possibilities for their abuse will become fully apparent only with extensive use—and misuse. Of course, one information user, through legitimate means, could gain the use of information in another organization's memory. Certainly, governmental interagency cooperation is inevitable. This is likely to occur also in some forms of interbusiness cooperation. As computers are interconnected more and more, as related organizations come to share specific memory banks, the inclination to share information and the ease of doing so will increase. After all, one of the advantages of the large-scale use of computers is the savings to be made by eliminating duplication of information and by standardizing information collecting forms and data retrieval languages.

So far we have speculated on the increased opportunities which the computer provides for invading privacy from "outside" the person. But there are trends which suggest that many people are likely to cooperate in exposing their previously private selves. Systematic exposure of the private self through questionnaires, interviews, and test taking is becoming steadily more widespread and probably more acceptable. The pressures toward—indeed, the attractiveness of—this kind of exposure are strong. In the first place, many Americans like to believe that getting ahead is a matter of ability and personality rather than luck or nepotism or some other kind of whimsey in those who hire and fire them, or in those who acknowledge or ignore them when opportunities arise. As the society grows more complex and the individual's sense of his ability to influence it in his own interest seems smaller, the tendency to depend for placement and advancement on what can be revealed about oneself which can be evidenced and acted on "scientifically" may well increase. This would be a natural extension of our dependence on the expert, for it will be the expert who will assess one's "true" value—at least one's economic value—by evaluating the private information one makes available about oneself. This response also will be a natural extension of our dependency on the machine, which in this case will help the expert or make the decisions itself about the value of the individual, impersonally but with great precision, on the basis of what it knows. And the machine will do so "privately": it will not blabber secrets to other machines or other people.

Complementary pressures from those who would use information about the private person are likely to be great. The real or imagined need to use people efficiently will increase as more organizations find themselves in the throes of complicated and disrupting reorganization, remodeling people and procedures to meet requirements imposed by the use of automation and computers. Thus executives and decision makers, responding to emotional and practical pressures, will try to squeeze the utmost from available personal information as clues to efficient job assignments. Increasingly, executives seem to seek security through technological intervention in the conduct of their activities. This tendency will be reinforced as more executives become the products of physical or social engineering backgrounds. As the behavioral scientists' predictive capabilities are increasingly recognized, and as business and Government become more professionalized and rationalized through the combined impact of the availability and capabilities of computers and the hyper-rational orientations of the personalities who tend these devices, management will seek more and more to learn all it can about the people it uses and about the people it serves, in order that its tasks may be more efficiently conducted in the public interest or in its own. This is, of course, simply a continuation of a well-rutted trend. For some years we have assumed or accepted that efficient Government, Government subject to as little internal disruption as possible, "requires" personnel selection on the basis of very private information about an applicant's sex life, family affairs, and early ideological enthusiasms. Since most new jobs in the past two decades have been Government or Government-related jobs, we can expect this trend to continue, the result being that more people than ever will be on file.

We have noted the likelihood that, as the years go on, the pressures will increase substantially to collect and use available data about the private individual, to enlarge the scope of that data, and to attempt to correlate and expand its implications and meaning. To some extent, each user of information will gather his own, but, in keeping with standards of efficiency, there will be efforts to gain access to data accumulated by others. For example, we can expect pressures to combine credit rating information with Government job application information, with school psychological testing information, and so on. In an enormously complex society, everyone may have something to gain by this process as well as something to lose.

One thing is clear: for a long time indeed no correlation of data will account fully for the personality being evaluated and interpreted. Whatever the person providing the information believes, and whatever those using the information believe, there still will be a truly private person left, undetected by the computer. This is not meant in any mystical sense. It will take much longer than the 20-year period we are dealing with to gain enough understanding of human beings truly to strip them of all the private self which they think they volunteer to expose. Nevertheless, in many situations, that information which can be processed through punchcards or memory tapes will be accepted as the important private profile of the individual. This will be so because, limited as the data may be in some abstract sense, this will be the information most conveniently available to the users for the assessment of the individual. And it will not be trivial or simple information. It will be impressive in its scope, and the computer will be impressive in the processing of it. Thus the users will choose to believe that this is the important part of the private life of the individual, and from the economic standpoint it may well be. Similarly, many of those supplying the information will come to believe they are revealing their private selves. In other words, that which will be valued and acted on as if it were the private individual will be that which can be tested and assessed in ways which can be recorded and manipulated by computers. Of course, not everyone will succumb to this bifurcation of self, but enough may do so to make it an important factor in our society. The result may be that many will feel they or others have no private lives. Others will feel that their "real" private lives are even more private because they are relatively more ignored—the computer won't be able to do anything with them. Thereby we shall have a new measure of privacy: that part of one's life which is defined as unimportant (or especially important) simply because the computers cannot deal with it.

No one using the output from a computer needs to know as much about the data fed into it as does the programmer. Without intimate and extensive understanding of the data and the uses to be made of it, the programs which determine how the computer operates, and hence the quality of its output, will be crude. On the other hand, executive decisions often depend less on knowledge of details than on overall grasp of the situation. As a result, the programmer often will be the person with potentially the most intimate knowledge of the private lives of those whose data is processed. This potentiality need not result in his having specific knowledge about specific people, since a programmer is unlikely ever to see the materials which are input to the computer whose processes he has arranged. But given his deeper understanding of how the data are being processed, what assumptions are made about the relationships among the data, what constraints must be put on the data in order for the computer to use it, it is entirely possible that the programmer may be called upon in difficult cases to enrich the executive's basis for decision making. In this way, the programmer may become privy to very private information about specific individuals. There may then arise a demand for programmers with ethical standards which now are not considered prerequisites to their trade. Inevitably, of course, there will be corruptibles among this group who will leak private information.

In another sense the programmer will become important for the preservation of privacy and freedom. The way he arranges the relationships in the information to be processed and the relative emphasis he gives to different items could result in distortions in the "history" of the person and, hence, in the implications of the data. In other words, the programmer could invent a private life. The question then arises of how the individual protects and asserts his own version of his private life over and against that defined by the computer. In the past, it has been possible to refer differences in present interpretations of past events to witnesses or paper records or photographs. Such records were public in that they were visually comprehensible. But records storage will

become ever more bulky and retrieving information will become an increasingly awkward and vexing task as the population increases and as the amount of information about each member of the population increases for the reasons we have discussed. The incentives to put this information into computer memories will thereby increase. But if history is recorded on tapes, in magnetic codes, and on molecular films, the definition of what was will become ever more dependent on how the machine has been programed and what it is able to retrieve from its memories. As respect for and dependence on the computer increases, it is likely that respect for and dependence on fragile and "ambiguous" paper records will decrease, lessening the ability of the individual to establish a past history different from that jointly provided by the programmer of raw data and the interpreter of processed data. There will be fewer opportunities to derive a public consensus on what the data "is," for there will be no public language in which the primary data will be recorded through which the public can verify the meanings and facts of the records. Robert Davis of the Systems Development Corp. compares this situation to the time when the Bible was interpreted to the illiterate, and what the Bible said and meant depended exclusively on what those who could read claimed it said and meant.

On the other hand, centralization of private information and its preservation in computer memories may decrease illegitimate leaks of that information. Those who will have access to personal history will see much more of it than was usually the case when it was contained in printed records, but fewer curious eyes will have knowledge of any part of the private history of the individual.

#### PERSONAL FREEDOM

Now let us look at a few possible confrontations between the freedom of the individual and the computer.

There is one form of technology tied to the computer which today increases freedom for some and which may in the future decrease it for others. This is the technique of telemetering information from tiny sensors and transmitters embedded in the human body. Right now, one form of these devices keeps recalcitrant hearts beating steadily. In a few years, in variations of already existing experimental devices, they will transmit information about subtle internal states through a computer to the doctor, continually or at any time he wishes. Clearly, the lives and liberty of people dependent on such support will be enhanced, for it will provide greater opportunity to move and to live than would be theirs if this information were not so continuously and directly available.

It is not impossible to imagine that parolees will check in and be monitored by transmitters embedded in their flesh, reporting their whereabouts in code and automatically as they pass receiving stations (perhaps like fireboxes) systematically deployed over the country as part of one computer-monitored network. Indeed, if they wish to be physically free, it is possible that whole classes of persons who represent some sort of potential threat to society or to themselves may be required to keep in touch in this way with the designated keepers of society.

It may seem farfetched to suggest that such people might walk the streets freely if their whereabouts and physiological states must be transmitted continually to a central computer. But two trends indicate that, at least for those who are emotionally disabled, this is not unlikely. We are now beginning to treat more and more criminals as sick people. We are beginning to commit them for psychiatric treatment rather than to jail. This treatment may have to continue indefinitely, since frequently a psychiatrist will not be prepared to certify that his patient will not commit the same kind of crime again (as is now required for sexual offenders under psychiatric treatment). At the same time, chemical and psychotherapeutic techniques for inducing tranquil emotional states are likely to improve. We may well reach the point where it will be permissible to allow some emotionally ill people the freedom of the streets, providing they are effectively "defused" through chemical agents. The task, then, for the computer-linked sensors would be to telemeter, not their emotional states, but simply the sufficiency of concentration of the chemical agent to insure an acceptable emotional state. When the chemical agent weakens to a predetermined point, that information would be telemetered via the embedded sensors to the computer, and

\* Interview with Robert Davis of Systems Development Corp., in Washington, D.C.

appropriate action could be taken. I am not prepared to speculate whether such a situation would increase or decrease the personal freedom of the emotionally ill person.

Already the computer is being used in conjunction with other technologies to retrieve information customarily stored in libraries. Doubtless, this use of the computer will expand greatly. Tied in with telephone lines and television cables, it will make it possible to gain access to vast areas of knowledge without leaving one's own local area. To the extent that knowledge increases the individual's opportunities for growth and effective mobility, we could say that such access at a distance will increase his freedom. This will be especially true for those who do not live near the conventional repositories of information.

It is sometimes suggested that the computer will bring back townhall democracy by making it possible for every voter to express his opinion at the time his representative needs it, merely be a pushbutton response to a teletyped or mass-media-transmitted request for the constituent's position on a given topic. The voter would gain more freedom to express himself, but that of the representative to act in terms of his own estimate of the best interests of the Nation or his district might be lessened by such ubiquitous and massive grassroots expression. If the representative were able to determine the voter's understanding of the issue in order to assess the meaning of his pushbutton vote, however, and if he had information on voters so that they could be clustered according to background, thereby allowing patterns of votes to be more fully interpreted, the representatives' ability to act in the combined interests of his various constituencies and the Nation at large could be increased. Such data also would give him a better basis for providing his constituents with the information they need in order to vote more intelligently. The biggest unknown would seem to be whether one could count on developing an intelligent and enlightened public, or whether computers used this way would simply increase the likelihood of representatives being swayed or dominated by a mass incapable of judging the meaning or implications of the complex issues it is asked to evaluate.

If it is worthwhile, the enormous capacity of computers can provide the basis for differentiating among many subpopulations. This capability could mean increased responsiveness on the part of data users and planners to the different social, psychological, and material needs of each of these populations. It could lead to more opportunities for individual expression, at least to the extent that the substance of individual expression is significantly differentiated by membership in various subgroups. In this sense, the computer could provide a greater opportunity for freedom than would be available in a large society which had to plan and operate in terms of overall averages rather than differentiated averages. But this capability for recognizing differences in populations and interest groups could also be applied to more detailed surveillance, causing a much greater loss of freedom than would result in a large population dealt with in overall averages. In the latter situation, the individual could more easily lose himself in the mass.

Behavior is internally mediated by the individual's history, personality, and physical capabilities, and externally by the constraints which environment imposes or the opportunities which it provides. Knowledge of external environmental characteristics increases the observer's ability to predict individual behavior. This is evidenced by the ability of police, military intelligence, or traffic control experts to predict, from knowledge of their environments, likely behavior of specified people or groups. Obviously, these predictions are improved greatly when added to knowledge of the environment is knowledge of the typical response of the individual to similar environments. The police capturing the criminal when he returns to the scene of the crime, the successful guerrilla ambush, the parent removing "temptation" from the child's environment, and the "lead item" sale, are all examples of how the discovery of patterns of correspondence between environment and individual behavior toward it can be used to predict and channel that behavior. Better predictions about behavior should be possible when external environment can be codified and defined in much greater detail by using computerized data and by using the computer to detect patterns of sequences and arrangements of individual acts through elaborate analysis and synthesis of the data. The resulting predictions could be used to alter an environment (inanimate or human) in order to provide more opportunities for alternative behavior to facilitate habitual behavior, or to inhibit or terminate behavior.

As usual, the consequences of environmental controls for freedom could go either way. Improved surveillance techniques would mean less crime, or, at any rate, less than there would be without such techniques. Less crime means more freedom and privacy, at least for the law abider. But the same techniques could be used against the law abider if his Government wanted to make, say, a routine security check in the interests of social stability.

When the application of computers requires that people change their behavior toward something familiar, they may well interpret this as an imposition on their freedom. This interpretation is in keeping with the belief long held by many that the machine is the chief threat to the spontaneity (freedom) of man. The recent furor over all-digit dialing demonstrates how seriously this threat is taken.<sup>5</sup> In the abstract, at least, one's freedom to dial long-distance numbers direct may be increased by this new system, and certainly it is not lessened compared to what it was when one used a mixture of letters and numbers. But obviously many people feel their freedom has been abridged because for them it seems easier to remember combinations of letters and numbers, and because this change symbolizes more mechanization and, thereby, a challenge to the freeman. Undoubtedly, there will be further "invasions" of this sort.

An important variant of this state of mind is found in responses to the nationwide computerized system which makes it possible for a cashier to determine quickly whether an unfamiliar person seeking to cash a check has a criminal record. Through this system (cashier to computer to police) a number of criminals have been apprehended while they waited for their check to be cashed. Abhorrence of the system and sympathy for the bum-check passer is a common—although, of course, not unanimous—response to descriptions of this system in action.

Apparently, in many minds there is combined a sense of "There but for the grace of God \* \* \*" and a realization that the inclination to violence and law-breaking which most of us harbor will be throttled more and more even in fantasy. For what is mere man against the implacable, all-seeing machine? The godlike omniscience of the computer essentially destroys his hope, and hence his freedom to fantasy, that he can get even unfairly with a society which he thinks has been unfair to him. If the computerized world of tomorrow produces the kinds of rationalized standards which increase one's frustration and inhibition, then certainly this invasion of one's right to hope (i.e., to fantasy antisocial success), will be interpreted as some kind of invasion of his personal freedom. If so, there most certainly will be an acceleration of a trend already underway: "Frustrate" the machines. In a spirit of desperation and vengeance people are bending punchcards, filling prepunched holes, and punching out additional ones. (Injunctions have already made it clear that this destruction of private property will not be tolerated, regardless of its contribution to the preservation of psychic property; the machine wins.) They are also overpaying, by one cent, computer-calculated and computer-processed bills and refusing to use postal ZIP codes.

Now, it may well be that existing law or future decisions and actions of courts and legislatures will enforce and elaborate present legal powers in order to conquer the threats to freedom and privacy on which we have speculated. But seldom is a law promulgated in anticipation of problems, especially when there are powerful interests which benefit from freedom to exploit. Moreover, as we have seen, in most cases there may be a potential or actual gain for freedom or privacy along with the loss. And as we well know, even existing laws protecting privacy and freedom are often difficult to apply ubiquitously and effectively. In the hothouse world of Washington, D.C., it is commonly believed that anyone who is anyone at all has had or is having his phone tapped by Government agents. Whether or not this is true, what is important is that people believe it is true, and they accept this situation feeling either that the Government has a right to such spying or that, even if it hasn't, they can do nothing about it. We are all well aware of the increasing pressures to enlarge the search and arrest powers of the police in the face of expanding urban crime. We know, too, that in some places some of these powers have been granted or their unsanctioned use tolerated. And what shall we conclude about such modifications of the law as that represented by an executive order which gave the House Committee on Un-American Activities for the period of the 88th Congress

<sup>5</sup> See, e.g., Time, July 13, 1962, p. 53; Washington Evening Star, Jan. 31, 1964, p. A12, col. 2.

the right of access to any income, excess-profits, estate, or gift tax records it wished concerning subversives?<sup>6</sup>

In the face of the increasingly complex tasks of maintaining, to say nothing of improving, urban society; in the face of general popular disagreement at best, and indifference at worst, about the proper conduct and protection of the individual in this alienated and splintered world; in the face of the special advantages the computer provides to increase information, command, and control—I would speculate that, at least for some time to come, these advantages will generally be considered more important than protection or preservation of those threatened aspects of freedom and privacy we have examined here.<sup>7</sup>

<sup>6</sup> Executive Order 11109, Fed. 11109, 28 Fed. Reg. 5351 (1963).

<sup>7</sup> I am indebted to David T. Bazelon, Robert Davis, and Patricia McMonigle for their suggestions during the preparation of this statement.

## APPENDIXES

### APPENDIX 1.—REPORT OF THE COMMITTEE ON THE PRESERVATION AND USE OF ECONOMIC DATA TO THE SOCIAL SCIENCE RESEARCH COUNCIL, APRIL 1965

#### MEMBERS OF THE COMMITTEE ON THE PRESERVATION AND USE OF ECONOMIC DATA

Richard Ruggles, chairman.  
Richard Miller, secretary.  
Edwin Kuh, Massachusetts Institute of Technology.  
Stanley Lebergott, Wesleyan University.  
Guy Orcutt, University of Wisconsin.  
Joseph Pechman, Brookings Institution.

#### SUMMARY OF REPORT

During the past four years the Committee on the Preservation and Use of Economic Data has met with a considerable number of Federal agencies concerned with the collection and use of data in machine readable form. The prime concern of the committee has been the development and preservation of data for use in economic research. Although considerable progress has been achieved in specific areas, the committee has concluded that three more general lines of action are required. Specifically, these are (1) the Federal Government should undertake the establishment of a Federal Data Center; (2) procedures should be established to insure the development and preservation of import data; and (3) research institutions and universities should develop an organization for coordinating their requests for economic data.

First, the committee urges that the Bureau of the Budget, in view of its responsibility for the Federal statistical program, immediately take steps to establish a Federal Data Center. Such a Federal Data Center should have the authority to obtain computer tapes and other machine readable data produced by all Federal agencies. It would have the function of providing data and service facilities so that within the proper safeguards concerning the disclosure of information both Federal agencies and users outside of the Government would have access to basic data. The Federal Data Center would require computer facilities, and it would need to be staffed with personnel capable of understanding the data problems in the various areas. In view of the importance of the Federal Data Center, the committee suggests that the Center be established with specific responsibilities for these functions and interagency authority to carry them out.

Second, the committee urges that the Office of Statistical Standards of the Bureau of the Budget place increased emphasis on the systematic preservation in usable form of important data prepared by those agencies engaging in statistical programs. In both the initial budget for statistical programs and the subsequent review of ongoing work, the Bureau of the Budget should see that provision is made for the development of computer tapes of important data, together with the supplementary material required for interpretation.

Third, the committee recommends that at an early date the Social Science Research Council convene representatives from research institutions and universities in order to develop an organization which can provide a clearinghouse and coordination of requests for data made by individual scholars from Federal agencies. In addition, such an organization would serve the Federal Government in an advisory capacity and provide a mechanism for the development of data tapes needed for research purposes in specific areas of economic research.



*The Background and History of the Committee*

In December 1959, the executive committee of the American Economic Association devoted a part of its annual meeting to the consideration of the preservation and use of data for economic research. The AEA Executive Committee recognized that research in the social sciences in general, and in the discipline of economics in particular, to an increasing extent requires large systematic collections of microdata for the formulation and testing of hypotheses, and that collections of microdata are research tools for the social scientist much in the same way that books and manuscripts are for the humanities and laboratories are for the scientist.

Although the AEA Executive Committee agreed that the problem of developing and preserving important bodies of microdata was extremely important, they concluded that it was not feasible for the American Economic Association to undertake an effort in this area, since the organizational structure of the association makes it difficult for subcommittees to carry out substantive work of this nature. The executive committee therefore recommended that the Social Science Research Council set up a Committee on the Preservation and Use of Economic Data to study this problem and undertake any program of action which it might deem desirable.

## CREATION OF THE SSRC COMMITTEE

Accordingly, in December 1960 the Social Science Research Council arranged an exploratory discussion of the problem of data preservation and use, drawing on social scientists from universities and the Federal Government. This discussion revealed that although there was a large area of common interest among the various social sciences, there were also wide differences in approach and areas of concern when any specific aspect of data preservation and use was being considered. It became obvious that the problem was too broad and diverse to be solved by any simple general solution, and that a more concerted and focused attack on specific parts of the problem would have to be undertaken.

For this reason, a small committee was then appointed to explore the problems arising in the field of economic data alone. This committee consisted of individuals who were acquainted with the statistical work being done within the Federal Government, and who were also doing research using large bodies of empirical data. At the outset the newly constituted Committee on the Preservation and Use of Economic Data recognized that in order to maximize the effectiveness of its effort it should concentrate its attention on those areas which would yield the most valuable research materials per unit of cost. By definition this immediately excluded information which was widely scattered throughout the Government or stored in bundles in warehouses, and it suggested that attention should be focused on those bodies of information which were currently available in machine readable form. The Committee was concerned with the archival problem, as well as the problem of making existing data currently available to research workers, and for this reason it proposed to examine the disposal policy of the various Federal agencies as well as to try to develop methods whereby machine readable data could be made directly available to scholars engaged in economic research.

## COMMITTEE ACTIVITIES, 1962-1964

During the 3 years from 1962 through 1964, the Committee undertook to study, on an agency-by-agency basis, the problem of providing access to specific bodies of information. Meetings were held with a considerable number of independent agencies in the Departments of Commerce, Labor, Treasury, Agriculture, Interior, and Health, Education, and Welfare. In addition, the Committee kept in close contact with the Bureau of the Budget and the National Archives. In some cases, arrangements were made to create and make available specific bodies of information, and substantial progress was made in developing awareness of the general problem of preservation and use of data by the Federal agencies. Representatives of the National Science Foundation attended many of these meetings.

Despite the progress which was achieved in specific areas, however, the Committee, at the end of 3 years' operation, concluded that some more general solution was required. Such a solution would require that the Federal Government develop (1) a systematic policy insuring the preservation of important data, and (2) mechanisms whereby data could be made available for research pur-

poses to universities and research institutions. In addition, the Committee also recognized that the universities and research institutions themselves should develop a more systematic and coordinated program of data development.

This report is intended to set forth the conclusions of the Committee with respect to the problems inherent in the preservation and use of economic data collected by the Federal Government. In addition, part II will consider the problem of data development facing universities and research institutions, and make recommendations as to steps which can be taken in this area.

*Part I. The Responsibilities of the Federal Government in the Preservation and Use of Economic Data*

## DECENTRALIZATION OF THE FEDERAL STATISTICAL SYSTEM

The statistical system of the Federal Government is highly decentralized. In contrast with many other countries, the United States does not have a central statistical office which is responsible for the recordkeeping of the Nation. Instead, each of the large number of administrative and regulatory agencies undertakes to provide much of the information which is required for its own operation. Thus, for example, the Bureau of Labor Statistics collects information on wholesale and retail prices, wage rates, employment, and a wide variety of other data relating to the role of labor in the economy. The Office of Business Economics provides data on the national income accounts, showing the progress of business activities and the functioning of the economy. The Federal Trade Commission and the Securities and Exchange Commission collect quarterly financial reports on manufacturing corporations. The Department of Health, Education, and Welfare collects the basic statistics on education and health.

In a great many instances, these statistics themselves are a byproduct of the regulatory process. Thus the Internal Revenue Service processes personal and business income tax returns and provides statistical tabulations of these returns which constitute a basic statistical source. Similarly, the Social Security Administration, in carrying out its administration of the social security program, has large bodies of information on wage and salary payments to individuals. However, one Federal agency, the Bureau of the Census, performs many of the functions normally undertaken by a central statistical office. The Census Bureau is responsible for comprehensive data on population, housing, agriculture, manufactures, retail and wholesale trade, transportation, and government bodies. This information provides other agencies with basic information about the American economy and its functioning. Thus, for example, census data provide much of the information behind the national income accounts and the detailed data about specific States and cities. To an increasing extent, the Census Bureau is undertaking special services and tabulations for other Government agencies. Certain tasks formerly undertaken by other agencies, such as the collection of foreign trade statistics and labor force surveys, have become a regular part of the census program.

## RESPONSIBILITIES OF THE OFFICE OF STATISTICAL STANDARDS

The Office of Statistical Standards of the Bureau of the Budget has the function of improving, developing, and coordinating Federal statistical services. There are two specific ways in which the Office of Statistical Standards can enforce coordination and maintenance of high statistical performance. First, all survey forms which are sent out by the Federal agencies for the collection of data must have the approval of the Office of Statistical Standards. Second, since it is a part of the Bureau of the Budget, the Office of Statistical Standards participates in the review of budget requests of the various agencies for statistical activities. Both of these instruments are important, but unfortunately the task of coordination is so great that it is difficult to insure the comparability of data among various Federal agencies.

## RESPONSIBILITIES OF NATIONAL ARCHIVES

The National Archives and Records Service has the responsibility for promoting improved current records, management, and disposal practices of Federal agencies, and for selecting, preserving, and making available to the Government and the public the permanently valuable noncurrent records of the Federal

Government. Before the advent of the computer, the National Archives were concerned primarily either with the basic original records or documents obtained by the Federal agencies, or with the analytic or statistical end products. The problems of intermediate worksheets and data in semiprocessed form were left largely to the discretion of the individual agencies involved. Thus, for example, with respect to the corporate tax records of the Internal Revenue Service, the National Archives has preserved in warehouses bales of tax returns filed by corporations going back to 1909. In addition, National Archives has also preserved the statistical tabulations of tax returns. With the development of machine readable data, however, it is becoming increasingly obvious that bodies of information in machine readable form which are intermediate between the original records obtained by a Federal agency and the final statistical tabulations may be more worth preserving than the original records themselves. There is a growing recognition by the National Archives of this fact. The committee was very much impressed by the active interest which the staff of the National Archives showed in this problem. However, again the problem is so vast that it may require completely new procedures and policies in the future.

#### IMPACT OF THE COMPUTER ON DATA PROCESSING

Data processing methods have undergone a systematic evolution which has had far-reaching implications for the Federal statistics system ever since the original punchcard equipment was introduced. Early computers were to some degree a logical extension of this punchcard equipment. Although the Univac Model I pioneered by the Bureau of the Census in the early 1950's represented a monumental step forward, it was only the modest beginning of what has turned out to be a completely new technology. Each succeeding generation of computers incorporates improvements in the size of memory, the speed of computation, and the density of data storage on tape such that the capacity and speed of operate have been increased many times over. By now the technological revolution has become so great that a reexamination of the organization of the Federal statistical system is urgently needed.

##### *Increase in efficiency*

From the outset, the computer, like other forms of automation, has reduced the amount of labor required in the processing of data. Before their introduction, a large organization of clerks and punchcard machine operators was needed to handle the huge volume of punchcards required for any substantial statistical operation. Sorting, tabulating, and computing were relatively lengthy processes. Even for minimal tabulations a great many steps were requested. It is true, of course, that the computer has made necessary the development of specialists who could write programs for data processing, but once a program is written and proved out, it can be used to process large masses of information rapidly and with a small staff.

##### *Reduction in processing time*

Equally important, the time required for data processing has also been substantially reduced. Operations which formerly took 7 to 8 months to carry out now have been reduced to a matter of weeks. In the processing of the 1960 population census, the time required for certain steps was reduced from several years to several months. This shortening of time has not only meant an increase in efficiency in terms of overhead and other fixed elements in the program, but it has also resulted in making important information available more promptly. This reduction of the timelag between the collection of information and its availability greatly affects the usefulness of the information.

##### *Improvement in data quality*

The computer has also made possible new kinds of analysis which could not have been done before because of the cost and time required to carry out the necessary computations. First, it has become possible to examine and edit much more carefully than was possible heretofore. Computers can "wash" the information, and find inconsistencies which would have gone unnoticed in hand editing. Editing instructions to test the reasonableness of the basic information can be built into the processing programs. Thus, in the case of census data for manufacturing establishments, the computer can spot errors in reporting wage bills and manhours by computing average hourly earnings. Where the resulting figures are outside a reasonable range, the original information can be questioned. Other kinds of inconsistencies can be tested in a similar way, and for

each individual report the computer can make literally thousands of tests to determine which figures are out of line and which specific items should be corrected. For some Federal agencies, the ability of the computer to make such consistency checks is very important. Thus the Internal Revenue Service uses computers to check the internal consistency of items contained in each individual tax form. Such an operation is basic to one of the major administrative functions of this agency, but before the introduction of the computer it was too expensive and time consuming to be feasible. In such uses, the computer is adding a new dimension to the work and increasing the overall efficiency of the agency.

##### *Data reduction and tabulation*

Even with quite sophisticated punchcard equipment the difficulty of handling large masses of information made it imperative to reduce the information as quickly as possible to a more manageable volume. In the past this generally led to the development of a given set of tabulations, which became the final form of the data and which were all that was available to prospective users. Under these circumstances, the primary focus of attention by the producers of data was on the final published form of tabulation which was to result from the data processing. Once these tabulations were finished, there was little or no thought of utilizing the original reports for alternative analyses, since the cost and time required for additional data processing were too great. With the dramatic reduction in cost and time which the computer has yielded, however, the focus of attention is shifting to the basic information. It is now possible to use the same basic data again and again for different analytic purposes. From the point of view of analysis, the original unaggregated microinformation offer greater potential than tabulations of a more aggregative nature. Where relationships of data inherent in the basic reporting unit are important, aggregate tabulations often hide more than they illuminate.

##### *New types of analysis*

The ability of the computer to carry out detailed and complex computations on great numbers of individual cases at very high speeds has made it possible to make types of analysis which are not feasible without it. For example, prior to the introduction of the computer, aggregated tabulations of individual tax returns were used to estimate the impact which proposed changes in the tax law might have on total tax revenue and on particular classes of taxpayers. With the introduction of the computer, however, it became possible to develop a much more reliable method. A sample of 100,000 tax returns was obtained, and a computer program developed to recompute each tax return individually according to the proposed revision of the law, and thus show for the sample as a whole the exact impact of the change. This method not only provides a cumulative measurement in terms of total tax revenue, but also permits an analysis of which classes of taxpayers are affected, and by how much.

##### *Cost of electronic data processing*

Despite the very marked increase in the amount of data processing being done in Federal statistical agencies as a result of the introduction of the computer, the cost of data processing has become a smaller percentage of the total cost of obtaining information, and it represents a very small fraction of this total cost. In many cases the cost of the field survey may account for as much as 95 percent of the total cost, and processing the data less than 5 percent.

##### *Data storage*

The problem of storing basic statistical information has also been greatly reduced. A computer tape today will hold information equivalent to over 100,000 punchcards, so that a relatively small number of tapes may contain information which formerly would have occupied a great deal of space. Before the development of efficient tape storage, past data could not be kept for long periods by Government agencies, since room had to be made for the continuous inflow of new punchcards. For the first time it has now become feasible to keep the original information in machine readable form at very low cost.

##### *Increased use of data*

An important aspect following upon computer development has been the increasing use of basic computer tapes by others than the agency collecting and processing the original information. Prior to the computer, when the focus of data processing was the production of tabulations which would satisfy all users,

Federal agencies often published massive detailed tabulations which could be used by groups outside the agency for a variety of statistical purposes. With technological advances in both computers and printing, even more massive detail is being produced. As one small example, in the IRS statistics of income series the report on individuals for 1960 consisted of 165 pages, and 233 pages for 1961: on business it was 192 pages for 1960-61 and 274 pages for 1961-62. Aside from the cumbersomeness of the sheer volume of printed material, users of statistical information are now finding that the published tabulations are costly to use and often are unsuited for particular analyses which they wish to make. Even where a specific tabulation is exactly in the form desired, the user may find it necessary to put the data back into machine readable form before he can manipulate them. For these reasons, there has been an increasing tendency for Federal agencies to supply outside users with computer tapes of information to avoid the expense of recording the data. To an increasing extent, Federal agencies are considering the preservation of and accessibility to computer tapes to be a direct substitute for printed publication to make more detailed tabulations available to research users of data. Statistical programs are no longer viewed simply as projects involving the gathering, processing, and disseminating of information. Instead Federal agencies are developing the ability to tap into a source of information at one or more points in the processing stage, where data are in the form (after editing but before too much aggregation) and on the medium of recording (magnetic tape, not original schedules or printed reports) which are needed.

#### *Interagency use of data*

The ability of the computer to handle and interrelate large bodies of information has encouraged different Federal agencies to bring together information which they collect on related economic units. The recent development by the Bureau of the Census of enterprise statistics is an example. This set of information was created by linking the establishment data collected by census with corporation tax data obtained by the Internal Revenue Service. Previously, given the costs of processing and storage, only already tabulated sets of information could be brought together, and in most cases it was impossible to reconcile different sets of related data precisely. A byproduct of interagency cooperation has been an improvement in the comparability of classification systems, techniques, and methodology. In order to collate data from different sources, Federal agencies have found it necessary to use identical classification systems and to treat similar cases in a uniform manner. Information required as a basis for major legislative and executive policy decisions necessitates drawing on many kinds of data. The increasing ability of the various Federal agencies to integrate their basic data at a primary level will provide more reliable and meaningful information for policy purposes.

#### CURRENT PROBLEMS OF THE FEDERAL STATISTICAL SYSTEM

Although the development of the computer has solved a great many problems in the processing and handling of data, these very advances have raised problems which were not serious before, and until these problems are faced, the Federal statistical system will not reach its full potential.

#### *Preservation of data*

One of the first problems raised by the development of the computer is the preservation policy of the different Federal agencies. As already indicated, the information collected by the Federal Government represents a large investment of human and material resources both on the part of the Government in obtaining the information and on the part of the respondents in providing it. Before computers were developed the preservation of most of this information was not feasible because of the high cost of storage and the impossibility of low-cost retrieval. Now that large volumes of basic data can be kept conveniently and inexpensively in the form of computer tapes and processed at low marginal cost, the question of what should be preserved must be faced as a matter of national policy. Under the present decentralized Federal statistical system, it is extremely difficult to maintain a coherent and consistent policy with respect to the preservation of machine readable data. The various agencies are primarily responsible for day-to-day operations, and cannot give high priority to long-run considerations. There is no adequate mechanism for insuring that these agencies are following optimal policies with respect to the preservation of important

information. In view of the large number of organizations involved, it is inevitable that unless the situation is regularly reviewed by some group within the Federal Government which considers this problem to be a major responsibility, a satisfactory solution cannot be achieved. For this reason, the committee urges that the Federal Government develop procedures and mechanisms for insuring the adequate preservation of important data produced in any of the Federal agencies.

#### *Data access*

The problem of access to information is a very real one. At the suggestion of the Committee on the Preservation and Use of Data, the Bureau of the Budget and the National Archives jointly undertook a survey of machine readable data held by various Government agencies. The survey covered some 20 agencies in the Departments of Agriculture, Labor, Interior, Treasury, Commerce, and Health, Education, and Welfare, and the Board of Governors of the Federal Reserve System. Over 600 major bodies of data were listed in this preliminary survey. These data are stored on approximately 100 million punchcards and 30,000 computer tapes. The decentralized nature of the Federal statistical system makes it extremely difficult for users outside the Government, and even in other Federal agencies, to find out what data exist on various topics and how to obtain access to them. Different agencies have completely different policies with respect to access, and an outsider must know precisely whom to contact with respect to each specific kind of information. Most Federal agencies process data as an activity which is ancillary to their primary responsibilities, and therefore they find it inconvenient and costly to respond to specific requests for information which would necessarily disrupt and delay their own work. Even in the case of agencies such as the Bureau of the Census where an effort is made to respond to legitimate requests for information, it is often difficult to fit outside requests involving data processing into a work program in which the various stages of processing censuses or surveys have been carefully scheduled and timed. In other words, the present Federal statistical system is primarily geared to the production and processing of information for immediate administrative use or publication. Thus the present organization of Federal statistical operations does not lend itself to optimal use of the vast amounts of existing information, despite the fact that this use could be achieved at low data processing cost.

#### *Development of usable data*

Another major problem arising from the lack of supervision and coordination of data preservation techniques is that even where important data are involved. Federal agencies often fail to develop clean edited tapes and to provide supporting information about the data contained on the tapes. Under present circumstances, such inadequacies are quite understandable. In the processing of basic information, operating agencies are mainly concerned with achieving the results necessary for specific tabulations or given computations. It is unavoidable in this process that substantial errors will be found. Sometimes these arise from transcription or classification errors, or from errors in programming. For the purposes of the operating agency, errors can often be patched up on an ad hoc basis, but doing so leaves the original tapes with the errors in them. It is usually quite possible to correct such errors when they are known to exist. On investigation, however, the committee found that due partly to the fact that it was not necessary for immediate purposes and partly to lack of proper budget allocation agencies often neglect this task, even though its cost would not amount to more than 4 or 5 percent of the total computation cost, and in most cases would be less than 1 or 2 percent of the total budget for the project. In addition, agencies often do not provide sufficient information on the layout, classifications, and definitions of data contained in a tape. As a result, even for the agency's own purposes it becomes very difficult to go back after a few years and make use of the information, unless it happens to be in the same format and classification system employed for current data processing. The turnover of personnel within Federal agencies often make it impossible to trace back precisely what was done in the original coding of the schedules or programming. In view of these circumstances, what is needed is some system which will insure that for important data all Federal agencies will provide clean, edited data with accompanying information describing layouts, coding, and programming, so that these tapes can be served by both the agency itself and by other groups. Given the presence of day-to-day business, the shortage of funds which often occurs at the end of a project, and the priority

of other major responsibilities, Federal agencies cannot be expected to devote the required effort to the development of clean data tapes unless some specific procedures are developed to insure this result.

#### *Disclosure*

In addition to the problem of physical access to data, there is another factor which may prevent the utilization of data in their original form. A considerable portion of the information collected from individuals and businesses is obtained with the understanding that such information will be considered confidential and will not be available to other Federal agencies or anyone else. Thus, for example, in the case of the Annual Survey of Manufacturers, the data reported on the activities of manufacturing establishments constitute a confidential report to the Census Bureau and are protected by law from use by such agencies as the Internal Revenue Service for checking tax returns, or even by Congress in its investigations. It is recognized by all concerned that Federal agencies should not violate the confidentiality of their data by making them available to outside research workers or other agencies. However, it is often possible to disguise the information in such a way that specific data cannot be traced to any individual respondent. For example, the Census Bureau in the last few years has made available a sample of information on 100,000 individual households, giving considerable detail about the age, education, income ownership, occupation, etc., of the individuals in the household. In this sample the omission of detailed geographic information makes it impossible to trace the data to any specific individual. By using a similar approach the Internal Revenue Service developed a sample of 100,000 personal income tax returns. As in the case of the Census sample, data on individuals were provided without disclosing information that could be traced to any particular individual. For other types of data, the problem is somewhat more difficult. Thus, for example, merely indicating the size and industry of a manufacturing plant may be enough to identify it and so constitute disclosure, even if no additional identifying information is given. In many instances, however, there are ways in which such information can be utilized without disclosure. Since for most research purposes it is not necessary to present information on individual cases in the final results, it is often possible for researchers to provide compute programs which can be used directly upon the basic data under the auspices of the Federal agency responsible. Again, however, few Federal agencies are in a position to take the time and trouble to fill out such individual requests, even in those cases where the research would be valuable and outside financing is available.

#### CONCLUSION

In summary, therefore, because of the decentralized nature of the Federal statistical system and the pressure of the primary functions of the agencies, neither outside scholars nor Federal agencies are able to utilize efficiently the large amount of information which has been obtained at public expense.

#### PROPOSAL FOR THE ESTABLISHMENT OF A FEDERAL DATA CENTER

For the reasons outlined above, the Committee on the Preservation and Use of Economic Data urges that a Federal Data Center be established by the Federal Government to preserve and make available to both Federal agencies and non-Government users basic statistical data originating in all Federal agencies.

#### *Need for interagency authority*

The first and most basic requirement of a Federal Data Center is that it should have the authority to obtain computer tapes produced by other Federal agencies. The exact timing of the receipt by the Federal Data Center of such tapes will differ from agency to agency, and will depend on the kind of information involved. As a general rule, however, the Federal Data Center should obtain copies of the data when a clean, edited tape of the basic information first becomes available. Fortunately, because of the nature of computer processing, duplicate copies of the basic computer tapes can be produced at low cost, so that both the agency concerned and the Federal Data Center can simultaneously have the basic information available to them. In this connection the Federal Data Center should keep track of statistical projects underway in the Federal Government and make sure in advance that the budget for each project includes the proper provision for making clean, edited tapes and providing the necessary accom-

panying information on classification and programing. It should be the task of the Federal Data Center to follow statistical projects and to see that the clean, edited tapes are made available within a reasonable period.

#### *Need for computer capability*

The Federal Data Center will require substantial computer capability if it is to provide access to information by outside users and by other Federal agencies. It is important that the Federal Data Center should not only furnish basic information but also, on a reimbursable basis, it should make production runs and furnish aggregated tapes or results to scholars so as to eliminate many problems of disclosure. In a great many instances the Federal Data Center will find it advisable to develop new tapes combining information from various bodies of material produced by different Federal agencies. For example, the very considerable interest in data on specific regions or cities by State and local governments for programs such as urban redevelopment, welfare, and education, makes it desirable to combine various kinds of information pertaining to a specific area on a systematic basis. In many cases such information about communities and their characteristics does not violate any disclosure rules. These data are useful not only for purposes of public policy but also to business groups interested in market research and in planning longrun investment. It is important that the Federal Data Center be staffed with computer analysts who are subject specialists so that they can understand the nature of the data with which they work and can anticipate the analytical problems of the agencies and research organizations that want to use the data.

#### *Need for service facilities*

A Federal Data Center would provide servicing facilities, so that Federal agencies and individuals could obtain specific information directly, and it should publish descriptions of the data available. In this sense the Federal Data Center would serve somewhat the same role as the Library of Congress, inasmuch as it would be responsible for providing a systematic and comprehensive coverage of the material available in its areas of competence. It would also, of course, be serving the same function in the statistical area as Archives now does in the area of basic records and documents. It would insure that the most useful information was preserved in a usable form, and that duplicative and unwanted data did not clog the system. Finally, the Federal Data Center would provide basic information about the American economy as a primary objective rather than as a byproduct of the administration or regulatory function.

#### *Need for new administrative arrangements*

Although the functions described above for the Federal Data Center are in part covered by the activities of existing Federal agencies, no single agency is currently combining all of these necessary functions. The Office of Statistical Standards of the Budget Bureau does have the responsibility for the supervision and coordination of Government statistical activities, but it is not an operating agency. The National Archives also has interagency authority, but it has not been involved in the field of data processing and does not as currently organized have the ability or authority to undertake the task of selecting, monitoring, and controlling machine-readable data on the scale required. Finally, other statistical agencies of the Federal Government have the ability to handle, process, and combine masses of statistical data in an imaginative and productive manner, but these agencies lack interagency authority to obtain each other's records. Furthermore, although these agencies have been making an effort to provide reasonable access to their data the fact remains that they have major responsibilities for collecting and processing basic information on a continuing basis, and these responsibilities, which have first priority, make it difficult for them to devote adequate attention to individual requests.

#### *Need for early and positive action*

In view of these considerations, the committee concludes that immediate action should be taken by the Federal Government to establish a Federal Data Center and to insure the orderly preservation of important data. The Bureau of the Budget has been given the responsibility of developing programs and issuing regulations and orders for the improved gathering, compiling, analyzing, publishing, and disseminating of statistical information for any purpose by the various agencies in the executive branch of the Federal Government (see sec. 1 of Executive Order 10253, June 11, 1951). The committee therefore urges that



the Bureau of the Budget immediately take steps to establish a Federal Data Center which would have the functions described above. It should be recognized, furthermore, that the nature of such a data center is so different from anything now in existence that it may well require additional legislative authority so that its responsibilities can be well defined and recognized by all Federal agencies. It is very important that the Federal Data Center be conceived as a new and independent function, rather than an extension of present activities by any single Federal agency which has major responsibilities of another kind. In the development of the Federal Data Center it is to be expected that the Bureau of the Budget would consult with the various Federal statistical agencies involved with policymaking groups within the Federal Government such as the Council of Economic Advisers, and with congressional groups such as the Joint Economic Committee, as well as with research institutions and universities. It is to be hoped that this planning and preparatory work can begin immediately.

In addition to the early development of a Federal Data Center, the committee urges that the Bureau of the Budget place increased emphasis on the systematic preservation of important data by those agencies engaging in statistical programs. Specifically, the Bureau of the Budget should see that funds are budgeted for the development of clean tapes of important data together with the supplementary material required for their interpretation. The subsequent review by the Bureau of the Budget of ongoing statistical programs should make certain that the important data are in fact preserved in usable form. These procedures will be necessary even after the Federal Data Center is established, and they can be initiated immediately.

Finally, as an emergency stopgap measure, the Bureau of the Budget should undertake a current evaluation of the preservation policies of the various Federal agencies and together with the agencies make a joint determination of what sets of data should be preserved, and in some cases how these data can be put into a more usable form. In connection with this, it is also suggested that the Federal Government undertake to collect and publish at regular intervals an inventory of machine readable data held by the various agencies.

## *Part II. The Role of Research Institutions and Universities in the Preservation and Use of Economic Data*

### USE OF DATA IN ECONOMIC RESEARCH

Economic research has undergone striking changes during the last decade, due mainly to the advent of the computer. However, the present organization of the profession and its lack of access to major data sources impose serious obstacles in the way of optimal use of this new research development.

#### *Research techniques and their development*

Prior to the development of the computer, empirical research in economics was largely confined to the use of aggregative economic data in fairly simple models. Price indexes, production indexes, national income accounting, and industry statistics were used not only as frameworks for classifying information, but also as a means of data reduction. The limited capability of economists to process information forces them to deal with aggregations, which often obscured interrelationships among basic variables. With the development of the computer, however, low-cost data processing has been made available to economists, and as a result for many types of economic problems research technology has undergone substantial change. Economists can now specify and develop sets of data which are tailored to the research which they are undertaking. They can also process large quantities of data on a case-by-case basis, so that complex interrelationships can be studied at a microeconomic level. The use of simulation techniques on a large scale makes it possible to test the sensitivity of models to different assumptions, and to variations in specific parameters. For the first time, it has become possible to make use of the large bodies of existing information, which can be quite powerful in testing as well as suggesting theoretical hypotheses.

#### *Research projects and individual research*

There are currently in research institutions and universities many research projects each of which involves a large number of scholars. The Brookings-SSRC model of the U.S. economy, the Harvard economic research project on input-output studies, the simulation studies at the University of Wisconsin, and the

research on consumer behavior at the University of Michigan are all examples of large-scale projects using large bodies of data processed on high-speed computers. Other institutions where computer facilities exist are also carrying out research of this type. The value and productivity of this research depend in large measure on the character and quality of data available. It is not only major research projects carried out by teams of scholars that have changed, however; individual research by specialized scholars working in a particular area has also been affected. In many universities and research institutions, there is no more than one economist for a given specialty, and for this reason he must do his research as an individual scholar. It is still true that many economists engage in research on an individual basis, but where before the computer the cost of processing data and making computations was beyond the resources available to the individual scholar, today this is not as true. The existence of bodies of data and the computer is extending the horizon of such scholars and is placing in their hands powerful research tools. An increasing number of substantial and valuable research projects is being undertaken because information is available on a highly disaggregated basis in machine readable form.

### ACCESS TO DATA BY ECONOMISTS

The use of the computer as a basic tool in empirical economic research does, of course, require that there exist bodies of suitable data in machine readable form. Without appropriate data, the economist with a computer would be in the same position as a biologist with a powerful microscope but no biological specimens. With limited or inferior data he will be constrained to results of limited usefulness or doubtful reliability.

#### *Large-scale research projects*

For the most part, large-scale economic research projects have a considerable advantage in obtaining the kind of information they need. However, even in these cases, the committee has found that the situation is far from satisfactory. Federal agencies are not organized to provide data, and therefore delays and administrative difficulties may make it impossible to obtain the desired information. The problem of disclosure of basic information poses additional difficulties, and Federal agencies may use these difficulties as a convenient excuse at times when they regard themselves as fully preoccupied with their own problems, although devices could be worked out to safeguard the confidentiality of the data. Where cooperation is required between two Federal agencies for the development of interrelated data, the difficulties are generally so great that research institutions hesitate to undertake the task.

#### *Individual research*

The problems facing the individual research worker are many times greater than those faced by large-scale projects. First, it is often quite difficult for an individual to find out what information exists and what form it is in. Second, making arrangements with Federal agencies often requires substantial time and effort, and usually agencies are not receptive to the individual scholar unless he is well known. The cost of having the Government prepare data in a form suitable for research purposes is very high indeed, because it must be done on a special ad hoc basis which disrupts the agency's operations. For these reasons the individual researcher is usually not in a position to obtain specially developed bodies of material. However, tapes of standard or multipurpose information specifically designed to be sold for research purposes can be developed. As one example, the 1-in-1,000 sample of the population census prepared by the Census Bureau has provided many universities and research institutions with a set of basic information which can be used in a large variety of research projects. Over the long run, the individual research scholar may have to come to depend upon such standard bodies of data much in the same way as he previously depended upon published tabulations.

### DATA ACCESS FROM THE POINT OF VIEW OF THE FEDERAL GOVERNMENT

As has already been indicated, the various agencies of the Federal Government have administrative and regulatory responsibilities which constitute their major functions, and the production of statistical information and the data underlying it is usually ancillary to these major functions. Demands for data by a large number of organizations, including not only research economists but



also State and local government groups, businesses, and other Federal agencies, often place a severe and unwanted burden upon data-processing facilities and the time and energy of specialized personnel. Even when such work is done on a reimbursable basis, limitations of staff due to overall personnel and budgetary considerations and the ability to hire people making the filling of special requests a burden. Outside requests for data are often uninformed, unreasonable, and in view of the Federal agency, not worth while. Few outsiders can know enough about the data, their nature and characteristics to make sensible requests, or to have a realistic appreciation of the analytic limitations which the data impose. As already indicated, the disclosure problem is formidable and causes considerable uneasiness on the part of the responsible people in the data-producing agencies, but it also may be very useful as a shield to protect them from the nuisance of dealing with individual requests. The problem, as seen by the Federal Government as a whole when contemplating a request for data, could be reduced if research workers asking for data could get together and coordinate their requests.

#### ECONOMIES OF SCALE AND THE NEED FOR COORDINATION

Many of the requests for basic information on a specific subject by different research scholars are duplicative. However, since each research project will be designed in somewhat different terms and has different objectives, it is inevitable that the independent requests for information will not be identical. On the other hand, it is also quite possible that, if careful consideration were given to the matter, general master tapes might be designed in specific areas which would meet the needs of a large number of research projects. One of the difficulties with published information is that different research workers want different types of aggregations and classifications. Since it is now possible to provide data on a disaggregated basis, these differences are no longer relevant, and it becomes necessary only to specify the basic items of information to be included in the body of data.

This basic similarity in the demand for information on a given topic implies that considerable economies of scale could be achieved by coordination. If a single master tape would fully satisfy the demands of each user, designing special tapes for each user would be unnecessary. For this reason, the committee has undertaken a preliminary survey of 10 major areas of economic data to see whether or not it would be possible to construct sets of such basic tapes in these areas. On the basis of this examination it is the committee's considered conclusion that this construction not only is feasible from the point of view of economic research needs and objectives, but also would go a long way toward improving access to major bodies of data for scholars, and toward reducing the costs and alleviating the burden placed on the Federal statistical agencies.

#### DEVELOPMENT OF TAPES FOR SPECIFIC RESEARCH AREAS

The committee circulated to a group of research scholars working in various areas copies of the preliminary inventory of machine readable data recently collected by the Bureau of the Budget and the National Archives. In a large number of cases, these scholars prepared suggestions as to bodies of data currently in existence, which should be developed and made available to universities and research institutions on a low-cost basis. Included among the suggested data files are some which are currently available to research scholars and which certainly should remain available. For example, the Bureau of the Census has developed a program of making available for purchase large bodies of unpublished data in the form of computer tapes. The Internal Revenue Service and the Bureau of Labor Statistics have also developed specific tapes for sale. However, even where unpublished material is available on computer tape, it is often not in a form which is directly useful to the research worker. Data reduction to prevent disclosure or to select a manageable sample of data may be necessary. It would also be very useful if the research community could be better informed about what tapes exist in the various Federal agencies, and the cost of obtaining them.

A number of different data characteristics which are important for research purposes have been mentioned by research workers. (1) They point out that sets of data that are continuous over time are particularly valuable. This is especially true where information relating to a specific reporting unit is obtained at regular intervals so that changes taking place at the micro level could be ob-

served. (2) Even where continuous reporting by individual units is not available, sets of information for different periods which permit cross section analysis are very useful for research purposes since they permit examination of changes in structural characteristics and behavior. (3) It is emphasized that sets of data covering a wide range of items for a single reporting unit are more valuable for many purposes than larger sets of information which report on a smaller number of variables. In a great many cases it is the interrelationships among variables at the individual reporting unit level that are important for research purposes. (4) Sets of information which it is possible to match with other kinds of information are particularly important, even where the information contained in such sets of data may be quite narrow. Thus, for example, if a set of data tapes includes a social security number or some other identifying characteristic which would permit matching with similarly identified collateral information from another source, the tape is that much broader in its coverage. (5) Many sets of data are useful not because they are in themselves unique bodies of specialized information, but because they are already in a machine language and are capable of being manipulated at low cost, so that it is often easier and cheaper to use them than to have recourse to data already in tabulated form.

#### PROPOSAL FOR AN ORGANIZATION ON ECONOMIC DATA

In view of the increasing importance and usefulness of machine readable data for economic research, the committee recommends that economic research institutions and universities develop an organization to coordinate the requests by research scholars for economic data and to aid the Federal Government in the development of data for research purposes.

#### Usefulness of Federal economic data

The inventory of machine readable data held by the various agencies of the Federal Government and the results of examination of the inventory by research scholars in various areas has convinced the committee of the potential usefulness of such information for economic research. Out of the 600 items listed in the inventory, over 75 bodies of data can be identified as of prime importance for general research in the 10 areas listed. There is no doubt that this list would be considerably expanded if the scholars consulted had had available more detail on the exact contents of the different bodies of data. Furthermore this preliminary inventory was far from complete in its coverage of Federal agencies. Since there is such a large body of highly useful data, therefore, the committee believes that research institutions and universities should encourage the Federal Government to undertake the establishment of a Federal Data Center.

#### Need for coordination

There is a substantial and growing demand from scholars in research institutions and universities for bodies of machine readable data held by the Federal Government. These demands are highly duplicative in nature, but completely uncoordinated. It is quite likely that in many instances it would be possible to obtain agreement from scholars working in a given research area as to what sets of information would be most useful if developed by the Federal Government. Such sets of information would satisfy the needs of many research analysts, so that Federal agencies would not be faced with many different requests. From the point of view of the community of research scholars, there would be considerable advantages in providing a clearinghouse for information concerning economic data, since it is so difficult for the individual research scholar to discover what information exists in the different Federal agencies, who should be contacted, and how problems relating to the confidentiality of data may be solved.

#### Need for data development

It is not sufficient, however, merely to provide a clearinghouse and to coordinate individual demands for data. In a great many instances the research community should take an active role in advising the Federal Government how to develop and exploit a given body of economic data. The existence of a body of information can often stimulate valuable research activity. Thus, for example, the 1-in-1,000 sample of the population census was not developed as a response to specific research demand by scholars outside of the Government, but rather it was developed by the Census Bureau because they recognized the potential worth of this type of data. Those scholars who were consulted about specific research areas, furthermore, emphasized the need to integrate the different

bodies of data collected by different agencies. Although the Federal Government is continually taking steps to improve the comparability of classifications used by the different agencies, the task of integrating bodies of data is still a formidable one requiring substantial effort. For these reasons, research institutions and universities should actively participate with the Federal Government in planning the development of economic data in specific areas.

#### *Need for a coordinating organization*

In view of these specific tasks facing economic research institutions and universities, it is important to establish some continuing organization. The committee recommends that institutions outside the Government which have a sizable staff engaged in economic research involving the computer processing of large bodies of machine readable data join to form a coordinating organization on economic data. The function of such an organization would include the servicing individual research requests for economic data by providing a clearinghouse and information about the availability of data. Second, the organization should undertake an active program of data development in conjunction with the Federal Government.

#### *Organizational structure*

Although the organization might have a larger membership which would draw on its clearinghouse and information services, it would be desirable to establish an executive committee so that periodic working meetings could be held to determine matters of policy. If there is to be continuity in the organization, furthermore, there will have to be a permanent secretariat which can function on a day-to-day basis. In view of the importance of the Federal Government as a data source, it is recommended that this secretariat be located in Washington. Finally, it is also recommended that the proposed organization develop working subcommittees of scholars concerned with specific subject matter areas so they can advise the Federal Government on data development and the establishment of procedures for coordinating demands for data.

#### *Need for early and positive action*

The committee urges that at an early date the Social Science Research Council convene representatives from research institutions and universities currently engaged in research projects involving the use of empirical information, in order to develop an organization which can coordinate requests for economic data. The group which is convened should give specific consideration to (1) how the research interests of all nonprofit research organizations and universities can be facilitated; (2) what kinds of services can be provided for nonprofit research institutions and universities; (3) what kinds of coordination are considered to be desirable; (4) how the proposed organization is to be established, staffed, and financed; (5) in what way the proposed organization can assist the Federal Government in the establishment of a Federal Data Center; and, (6) in what way the proposed organization can provide the Federal Government with advice concerning the preservation and development of basic data.

The formation of a coordinating organization should not, however, be delayed until solutions are found to all of these questions. There is an urgent need for an organized group with staff support to follow through on the problems outlined in this report. Such a group would be useful to the Bureau of the Budget in carrying out the suggestions contained in part I of this report. Further delay may result in the loss of valuable data which could be saved by prompt action. Furthermore, in order to provide for the orderly flow of data in its most useful form 2 or 3 years hence, steps must be taken now to establish procedures for projects which are already in their formative stages and which, unless properly conceived, may in 2 or 3 years time present the same sort of problems which are now encountered. Finally, the very rapid growth of research needs and the large quantity of machine-readable data generated tend to produce a large number of ad hoc solutions which will make future coordination more difficult. Adequate consideration of how to meet the needs of various groups in the immediate future may forestall the development of inappropriate partial solutions.

## APPENDIX

### PRELIMINARY EVALUATION OF PUNCHCARD AND COMPUTER TAPES OF ECONOMIC DATA HELD BY FEDERAL AGENCIES

Although it has not been possible to compile a comprehensive listing of the comments received from research scholars, the committee has made a summary listing of some of the punchcards and tapes mentioned by research workers in specific areas. In some cases, the items discussed include tapes which can be purchased from Government agencies as well as those which are currently unavailable. In a few cases, sets of information not included in the inventory of tapes were also mentioned. The numbers in parentheses which are cited refer to the inventory numbers given in the appended listing.

#### 1. POPULATION

The census of population data are not only basic to the study of demography, but also provide valuable information on individuals and households necessary for research on housing, employment, education, health, and consumer behavior.

##### *Census of population samples*

The Bureau of the Census has prepared 1:1,000 and 1:10,000 samples of the census of population (41-A-12 and 41-A-13) on both punchcards and computer tapes which are available for purchase. These bodies of data were warmly received by the profession and many research centers have purchased these sets of data. Many Ph. D. theses, as well as other research projects, are using this sample.

##### *Vital statistics*

The Public Health Service provides annual statistics on births, deaths, marriages, and divorces (68-12, 68-13, 68-14, 68-15, and 68-16), which are all very useful for simulation models involving population projections.

#### 2. HOUSING AND REAL ESTATE

Data relevant to research on housing and real estate are included in the material discussed under the headings of population, consumer behavior, agriculture, banking, and taxes. In addition, however, specific housing information is collected by a number of Federal agencies. Some of these agencies, such as the Federal Housing Administration, the Federal Deposit Insurance Corporation, and the Federal Home Bank Board, were not included in the inventory of machine-readable data and so are not referred to in this evaluation.

##### *Inventory of housing*

The Bureau of the Census provides a number of different series relating to this topic. The survey of inventory change and residential financing of housing units (41-A-22), the housing vacancy survey (41-A-16), and the housing sales survey (41-A-151) are all very useful, but for many purposes some data selection might be required to reduce the number of tapes.

##### *Building permits*

A number of different census surveys are available on building permits. Building permits issued monthly and annually (41-A-148 and 41-A-152), building permits used (41-A-149), nonpermit construction starts (41-A-147), and the construction progress report (41-A-150) are all relevant and important for the analysis of the construction industry.

##### *Prices of housing*

The Bureau of Labor Statistics price data on housing include consumer price data on housing (44-B-4) and rents (44-B-5). These tapes are in addition to the information on consumer expenditures included under consumer behavior, and if available in regional detail, would provide valuable information on the relative demand and supply of housing.

*Farm real estate*

The Department of Agriculture provides information on farm real estate values (40-F-10). In addition there are two surveys, farm real estate market survey (40-F-12), and farm building survey (40-F-14), which would be very useful.

*Other data*

Financial information provided by the FHA series on insured home mortgage terminations (55-17), and data on individual parcels given by the census of governments assessed valuations (41-A-46), would be particularly valuable. Although the "County and City Data Book" is published, the computer tapes of this information which are now sold by the Census Bureau are very useful for research.

## 3. LABOR FORCE AND WAGES

Information on the labor force, employment, earnings, and labor unions are provided by several Federal agencies. These data are often needed in a highly disaggregated form so that they can be related at a detailed level to other regional, industrial, and demographic information.

*Labor force*

The basic data in this area is provided by the Census Bureau in the Current Population Survey (41-A-19), and high priority should be given to making this available. In addition, the Social Security Administration provides useful samples of employer-employee records (72-1), and continuous work histories (72-3 and 72-4). Some sample of the summary earnings record tape (72-6) would also be desirable. Finally, the Bureau of Employment Security of the Department of Labor gives data on the employment and wages of workers covered by unemployment insurance (44-A-8), labor turnover (44-B-28), and the characteristics of the insured unemployed (44-A-9). Given the current interest in the problem of unemployment data these sources are very important.

*Wages and hours*

The Bureau of Labor Statistics collects the basic information in this area in its survey of industry employment, payrolls, and hours (44-B-27 and 44-A-18). The data on wages and related benefits for 82 market areas (44-B-14) also are highly important. Since census data on industry and trade also contain employment and wage data it will often be found useful if these various bodies of data are available in a form that can be interrelated.

*Unions and pension plans*

The characteristics of labor unions (44-D-12) together with their financial data (44-D-13 and 44-D-14) provide the basic information on labor unions. Pension and welfare plans are covered by additional Bureau of Labor Statistics surveys (44-D-6, 44-D-7, 44-D-8, and 44-D-9). The growing importance of pension and welfare funds both as a source of funds in the economy and in terms of effects on the future income of the aged make this information particularly valuable.

## 4. EDUCATION

The increased interest in education and the magnitude of expenditures on education make it imperative that adequate data on this topic be available for research purposes. Much of the basic information is contained in the population census and other surveys where data are provided on the age, sex, and educational attainment of individuals. However, the Office of Education of the Department of Health, Education, and Welfare, provides a considerable amount of specialized information.

*Primary and secondary schools*

The inventory of schools for resource evaluation (51-4) provides basic data on primary and secondary schools. Additional surveys of nonpublic schools (51-7, 51-18, and 51-19) are carried out on a periodic basis. Expenditures by type per pupil (51-9) and data on various aspects of the curriculum such as science and mathematics (51-3 and 51-8) and foreign languages (51-39) furnish valuable information on the extent of educational benefits in different areas.

*Higher education*

There is a considerable body of information available for colleges and universities in machine readable form. Data on plant and equipment (51-16, 51-20), enrollment (51-10, 51-11, 51-15), residence and migration of students (51-12), earned degrees (51-13), faculty (51-14, 51-34), and financial statistics (51-24), are available. The survey of scientific and technical personnel (41-B-49) made by the Bureau of the Census is pertinent here. All these bodies of information are important to research projects on the role and development of higher education in the Nation.

## 5. HEALTH

In view of the development of both private and public health plans, economic research on health has become very important. The Public Health Service has since 1959 provided a series of continuing surveys and a number of special purpose supplements, all aimed at establishing basic and comprehensive data for research in the health field.

*Health interview survey*

This survey (68-1) together with the personal health expenditure survey (68-9) should be made freely available to research workers with proper measures developed to safeguard the confidentiality of the original records.

## 6. CONSUMER BEHAVIOR

The field of consumer behavior has been intensively studied by economists for several decades. Consumer expenditure studies, analyses of purchasing intentions, and the financial characteristics of households are all important.

*Consumer expenditure*

The Bureau of Labor Statistics survey of urban consumers (44-B-6) and the Department of Agriculture survey of rural consumers (40-C-2) constitute the most recent basic data in this area. The committee has already indicated that these sets of data should be available for research purposes.

*Purchasing intentions*

The quarterly survey of the intentions of households collected by the Census Bureau (41-A-18) constitutes a body of information which is very useful in the study of consumer behavior.

*Financial characteristics of households*

The 1963 survey of financial characteristics (55-1) was made by the Bureau of the Census for the Federal Reserve Board. These data are valuable for research not only on consumer behavior, but also on the role of the household sector as a source of financing in the economy.

## 7. AGRICULTURE

Agricultural economic research has for many decades been a major concern of many colleges and universities in the United States. It is difficult at this juncture to specify just what categories of data would be of particular interest to the various research groups in these institutions. Nevertheless it is apparent that questions of land use, conservation, productivity, farm management, and many other topics are very important.

The inventory of machine-readable data in the Department of Agriculture covers six areas: (1) Forest Service, (2) Commodity Exchange Authority, (3) Statistical Reporting Service, (4) Agricultural Stabilization and Conservation Service, (5) Agricultural Marketing Service, and (6) Economic Research Service. There are, of course, several classes of users for this information. On the one hand, there are research groups interested in the economic conditions in agriculture within specific regions of the country, and for these groups highly detailed information of a sample nature is often very useful. Other groups are more interested in the total national picture, and the functioning of agriculture as a sector in the economy. These groups want comprehensive tabulations, some of which are supplied by the Bureau of the Census.

*Forest Service*

The Forest Surveys are often based on a two-stage sampling scheme using aerial photographs. They are of interest primarily to those analyzing regional

forest problems. Over 20 of these surveys were listed by the research workers consulted as having considerable priority. With respect to larger bodies of data, the National Compilation of Forest Survey Statistics (40-A-59) is a very important set of data, containing information on ownership, size, forest type, species, and timber products.

#### *Agricultural production*

The basic information on farms, farm characteristics, livestock products, crops, fruit, etc., is provided by the Census of Agriculture (41-A-6 and 40-F-57). The Department of Agriculture also has tapes on the June-December enumerative survey, providing acreage reports for crops and reports on livestock (40-C-12). Both of these sets of data are important in the analysis of agricultural output.

#### *Agricultural marketing*

The data provided by the Commodity Exchange Authority showing futures transactions and trading data (40-B-3 and 40-B-5) are useful in market pricing studies. In addition, some of the data provided by the Agricultural Marketing Service on such things as fruits and vegetables (40-E-3, 40-E-5), slaughtering (40-E-8), and milk (40-E-10 and 40-E-11) provide information on specific commodities in considerable detail.

#### *Farm management*

The Economic Research Service of the Department of Agriculture provides a great deal of information about the status of farmers and farm management. Over 30 sets of data were listed as being particularly important for research on such topics as the financial condition of the farmer, transportation, housing, real estate and land use. In addition, gross income, cost of production, machinery costs, and fertilizer costs and benefits are all topics of research interest for which important sets of data exist.

### 8. BUSINESS AND INDUSTRY

As already indicated, reports on specific business or industrial establishments might result in disclosure of confidential information. However, highly disaggregated data for regions and industries can often be presented without disclosure. In addition, samples may be developed which would not violate confidentiality, and fuller and more detailed data could be kept in a similar form by government agencies for those research projects which require processing of the original reports.

#### *Manufacturing and mineral industries*

The census of manufactures and mineral industries for 1947, 1954, and 1958 (41-A-32, 41-A-33, and 41-A-34) and the annual survey of manufactures (41-A-38) should be made available in as disaggregated a form as the disclosure rules will permit, and specific samples of data should be integrated with the census of manufactures data and the Internal Revenue data to provide more comprehensive and complete coverage of the manufacturing and mineral industries.

#### *Trade and services*

The economic censuses of wholesale (41-A-36) and retail trade (41-A-45), transportation (41-A-37) and services (41-A-42) should be treated in a manner similar to that described for data on manufacturing and mineral industries. In addition, the monthly surveys in this area (e.g., 41-A-51 to 41-A-60) should be developed into systematic sets of samples available over time.

#### *Banking and finance*

In the preliminary inventory of machine readable data in the Federal Government, the Federal Reserve Board was the only financial institution included. It is probable that when the survey is extended to other Federal financial institutions, many important bodies of data will come to light. In the material examined in the current inventory, member bank loans to commercial and industrial borrowers (55-49 and 55-50) and small business financing experience (55-45 and 55-46) obtained by the Federal Reserve Board represent valuable research materials for analyzing business financing.

### 9. GOVERNMENT FINANCES AND TAXATION

A large body of information is available on the income and expenditures of Federal, State, and local governments, and Federal tax returns of individuals and business. These data are a very valuable source of research material.

#### *Government finances*

The Census of Governments financial data (41-A-47) provides information on about 91,000 government bodies. These data are valuable for comparative research on State and local governments, and when used together with other regional information provide material for analysis of regions and standard metropolitan areas.

#### *Taxes*

The Internal Revenue Service and the Brookings Institution have created in recent years tax models for individuals (48-11), for corporations (48-12), and for partnerships (48-13). All these tax models have been found to be extremely useful for research purposes. Additional tapes have also been prepared for fiduciary returns (48-9), estate taxes (48-8), and gift taxes (48-10). It is recommended that tapes be prepared on returns showing capital gains and losses, and that continuous income histories covering both individuals and corporations be developed. Finally, a considerable number of scholars consulted by the committee emphasized that it would be highly productive if the tax records could be matched with social security records, the census of population, the census of manufactures, and the financial reports of the Federal Trade Commission and the Securities and Exchange Commission.

### 10. FOREIGN TRADE AND PAYMENTS

Considerable progress has been made over the last decade in the development of data on foreign trade. Imports and exports on a commodity and country basis are available in considerable detail on a monthly and an annual basis.

#### *Exports and imports*

The Bureau of the Census processes the basic foreign trade data. Export and import data are available for both waterborne and airborne trade for various levels of commodity and country detail. Such data are useful for a wide variety of purposes.

#### *Capital flows*

Analysis of direct foreign investment and short-term capital flows is important for understanding the balance of payments of the United States. Some of the required data now obtained by the Treasury, the Federal Reserve Board, and the Department of Commerce are in machine-readable form. However, a great deal of other important data still are not in this form. A significant contribution would be made as a first step by putting all balance-of-payments material on tape.

### 11. OTHER AREAS

There are, of course, other areas of data which deserve special attention. For example, data on the natural resources of the United States are very important for studies of conservation and research on the future growth of the economy. Special topics such as research and development expenditures by government and industry, studies of pollution, highways, railways, and impacted defense areas all require and often produce special sets of data which should not be overlooked.

#### EXCERPTS FROM AN INVENTORY OF PUNCHCARDS AND COMPUTER TAPES HELD BY FEDERAL AGENCIES DECEMBER 1964

This listing represents an initial inventory of the machine-readable data held by a number of Federal agencies. It was initiated by the Office of Statistical Standards of the Bureau of the Budget, with the cooperation of the National Archives and Records Services. The purpose of this inventory is to provide information for development of policies relating to the preservation and use of data collected and processed by Federal agencies. This summary of the inventory was prepared by the Committee on the Preservation and Use of Economic Data of the Social Science Research Council.



## DEPARTMENT OF AGRICULTURE

## FOREST SERVICE

- 40-A-1 Illinois Forest Survey (decennial): Number of trees, volume, and growth in cubic feet and board feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1962; two tapes.
- 40-A-2 Missouri Forest Survey (decennial): Number of trees, volume, and growth in cubic feet and board feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1959; 80,000 cards.
- 40-A-3 Minnesota Forest Survey (decennial): Number of trees, volume, and growth in cubic feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1960-62; 130,000 cards; three tapes.
- 40-A-4 Kentucky Timber Cut (decennial): Timber products, output, and resulting timber cut reported by product in standard units of measure. Unrestricted; 10 years; 1962; 500 cards.
- 40-A-5 Pulpwood Production in Lake States Counties (annual): Pulpwood receipts by quantity, source, and species, reported by primary wood using plants (pulpmills). Confidential; 2 years; 1,500 cards.
- 40-A-6 Small Forest Ownership in Southern Michigan (one-time survey): Social and economic characteristics of woodland owners. Reporting unit: woodland owners. Confidential; 7-8 years; 1959.
- 40-A-7 Recreational Use of Huron-Manistee National Forest (one-time survey). Purpose of visit, type of area, likes and dislikes, length of stay, etc. Reporting unit: Recreational groups. Unrestricted; 3 years; 1962; 700 cards.
- 40-A-8 Boundary Waters Canoe Area Recreation Study (one-time survey): Length of stay, purpose of visit, activities, likes, and dislikes. Reporting unit: recreational party. Unrestricted; 3 years; 1960-61; 4,214 cards.
- 40-A-9 Survey of Campers in Huron-Manistee National Forests (one-time survey): Income, age, occupation, education, family size, residence, amount of camping. Reporting unit: family campers and camping groups. Unrestricted; 3 years; 1962; 500 cards.
- 40-A-10 Fire Statistics (annual): Information on individual fires by cause, size class, fuel type, etc., man-hours of suppression action, type of action, cost of damages, etc. Reporting unit: Ranger district. Unrestricted; permanent; 1940 to present; 39,000 cards.
- 40-A-11 Timber Cut and Sold (quarterly): Volume and values of timber sold and cut by sale size class and species group. Reporting unit: National forest. Unrestricted; 5 years; 1959 to present; 9,000 cards.
- 40-A-12 Mill Scale Studies (selected intervals): Volume, quality, and value of units of output (boards and veneer) per log for a tree species. Reporting unit: processing plants. Unrestricted; permanent; 1954 to present; 275,000 cards.
- 40-A-13 Timber Inventory (decennial): Inventory of timber volumes, rate of tree growth, and tree mortality. Reporting unit: National Forest (working circle). Unrestricted; 10 years; 1954 to present; 700,000 cards.
- 40-A-14 Butte County, Fire Prevention Survey (one-time survey): The forest knowledge level (all aspects) of an individual. Reporting unit: an individual who is considered representative for a segment of the California population. Confidential; permanent; 1963; 300 cards.
- 40-A-15 Fire Weather and Fire Indices (daily): Wind speed, direction, temperature, humidity, precipitation, fuel moisture. Reporting unit: 335 stations run by DCF and FS. Unrestricted; permanent; 1951 to present; 80,000 cards; 190 tapes.
- 40-A-16 Survey of Timber Cut and Timber Products Output (one-time survey): Volume cut by species, by county of origin, by product, by ownership classes. Reporting unit: Wood using firms and operators. Confidential; 5 years; 1962; 12,000 cards.

- 40-A-17 Forest Survey, Hawaii (annual): Location, ownership, condition of forest land, volume, kind and quality of timber trees, the net annual growth of commercial timber, mortality, losses, and annual timber cut. Reporting unit: individual forest survey plots. Unrestricted; permanent; 1958-61; 53,000 cards.
- 40-A-18 Forest Survey, California (annual): Location, ownership, condition of forest land, volume, kind and quality of timber trees, the net annual growth of commercial timber, mortality, losses, and annual timber cut. Reporting unit: individual forest survey plots. Unrestricted; permanent; 1958-62; 200,000 cards.
- 40-A-19 Level and Sources of Fire Prevention Knowledge of California Hunters (one-time survey): Forest knowledge level (all aspects) of each hunter. Reporting unit: individual licensed hunter. Confidential; permanent; 1959-60; 1,400 cards.
- 40-A-20 Forest Products Marketing Research, Region 5, Timber Sales 10 (annual): Cost and revenue of selected processing plants. Reporting unit: individual firms. Confidential; permanent; 1951-61; 5,000 cards.
- 40-A-21 Forest Products Marketing Research, Region 5, Timber Sales 01 (annual): Characteristics of individual timber sales, date, seller, number of bidders, locations, etc. Reporting unit: individual sales. Unrestricted; permanent; 1952-61; 3,000 cards.
- 40-A-22 Information on Campground Use and Visitor Characteristics (daily): Number of persons per group, length of stay, place of residence. Reporting unit: visitor group. Unrestricted; permanent; 1958-61; 5,000 cards.
- 40-A-23 Campground Attendance, California (daily): Number of persons per group, length of stay, place of residence. Reporting unit: visitor group. Unrestricted; permanent; 1961-62; 500 cards.
- 40-A-24 Snow Course Water Equivalent (selected intervals): Water equivalent, average of five or more points, course elevation, slope, aspect, forest density, and associated meteorological conditions at index station. Unrestricted; permanent; 1958-60; 2,000 cards.
- 40-A-25 Watershed Characteristics and Conditions, California (one-time survey): Area-elevation slopes, aspects, stream lengths, geologic rock type, soil characteristics, area burned, vegetation cover types, and densities, other land uses (roads, logged areas). Reporting unit: individual watersheds. Unrestricted; permanent; 1955-64; 140,000 cards.
- 40-A-26 Individual Fire Reports (annual): Size of fire, cause, discovery time, attack time, control time, manpower, fuel type, suppression costs, and damages for bigger fires. Reporting unit: individual fire. Unrestricted; permanent; 1940 to present; 70,000 cards.
- 40-A-27 Lumber Manufacturing Costs and Selling Values (annual): Cost of manufacturing lumber and the selling value of it. Reporting unit: lumber mills, region 1. Confidential; permanent; 1962 to present; 50,000 cards.
- 40-A-28 Mill Scale Study, Region 1 (one-time survey): Quantity of board recovery from trees. Reporting unit: individual trees. Unrestricted; permanent; 1961-64; 400,000 cards.
- 40-A-29 Resource Accounting, Region 4 (one-time survey): Individual tree stands, acreage, species, site quality. Reporting unit: individual tree stands, 5 acres and larger. Unrestricted; 10 years; 1958-59; 10,000 cards.
- 40-A-30 Forest Fire Research, Region 4 (annual): Causes of fires, size and cost of fires by classes. Reporting unit: individual fire. Unrestricted; permanent; 1942 to present; 22,000 cards.
- 40-A-31 Recreation Facilities, Region 4 (annual): Inventory of recreation sites by classes. Reporting unit: individual campground site. Unrestricted; permanent; 6,000 cards.
- 40-A-32 Timber Inventory, Region 4 (one-time survey): Identifies volume of timber by species as merchantable or nonmerchantable. Reporting unit: geographic area (working circle). Unrestricted; permanent; 35,000 cards.
- 40-A-33 Wilderness-Use Study, Pacific Northwest (one-time survey): Name and address of registrant, number in party 16 years plus and minus, mode of travel. Reporting unit: recreational parties on wilderness trails. Confidential; permanent; 1961; 4,600 cards.



- 40-A-34 Wilderness-Use Study (one-time survey) : Name and address of registrant, number in party over and under 12 years; mode of travel, length of stay. Reporting unit: recreational parties on wilderness trails. Confidential; permanent; 1962; 5,000 cards.
- 40-A-35 Coast Douglas-fir Tree and Log, Lumber and Veneer Recovery (one-time survey) : Tree and log characteristics, lumber and veneer grade yields. Reporting unit: sawmills and veneer mills. Unrestricted; permanent; 1964-66; 200 cards.
- 40-A-36 Western Larch Tree and Log Lumber Recovery (one-time survey) : Tree and log characteristics, lumber grade yields. Reporting unit: individual logs. Unrestricted; permanent; 1961-62; 40 cards.
- 40-A-37 Inland Douglas-fir Tree and Log Lumber Recovery (one-time survey) : Tree and log characteristics, lumber grade yields. Reporting unit: individual logs. Unrestricted; permanent; 1961-62; 150 cards; 5 tapes.
- 40-A-38 Dwarfmistletoe Growth Impact Study (one-time survey) : Tree heights, age, d.h.h., vigor classification, disease rating, site quality, and decadal radii back to 1860 (decadal volumes were computed and are on second set of cards). Reporting unit: individual tree. Unrestricted; 10 years; 1960; 2,100 cards.
- 40-A-39 Forest Employment Data, Washington and Oregon (annual) : Number of employees in forest industries. Reporting unit: four-digit industry. Unrestricted; permanent; 1936 to present; 8,000 cards.
- 40-A-40 Forest Inventory Data and Related Inventory Studies (annual) : Forest resource statistics on area, volume, growth and drain. Reporting unit: forest inventory plots. Confidential; permanent; 1,100,000 cards.
- 40-A-41 Timber Growth and Growing Stock Projections (TRAS-2) (decennial) : Forest inventory, growth, mortality and drain. Reporting unit: data collected on forest inventory plots. Forest Survey section, for PNW Forest and Range Experiment Station. Unrestricted; 5 years; 2,000 cards.
- 40-A-42 Continuous Forest Inventory Data Cards, Region 9 (annual) : Annual survey of several local woods covering tree growth, mortality, insect and disease affliction, etc. Reporting unit: sample plots. Unrestricted; permanent; 1945 to present; 60,000 cards.
- 40-A-43 Forest Inventory and Utilization Statistics for Alaska (annual) : Furnish statistics on forest land areas, timber volumes, growth, mortality, quality, are condition, timber cut, and utilization practices. Reporting unit: trees and plots. Unrestricted; permanent; 1954 to present; 70,000 cards.
- 40-A-44 Black Cherry Lumber Grade Recovery Study, Northeastern Region (one-time survey) : Classification and measurement of lumber cut from individual logs; i.e., width, length and grade. Reporting unit: individual board. Unrestricted; 5 years; 1960; 12,000 cards.
- 40-A-45 Forest Survey of Pennsylvania (one-time survey) : Tree measurements on sample plots randomly located, diameter, height, species and grade. Reporting unit: individual trees in sample plots. Unrestricted; permanent; 1963-64; 45,000 cards.
- 40-A-46 Forest Survey of Maryland (one-time survey) : Tree measurements on sample plots randomly located: species, diameter, height and grade. Reporting unit: individual trees in sample plots. Unrestricted; permanent; 1962-63; 24,000 cards.
- 40-A-47 Forest Survey—Alabama, Arkansas, Louisiana, Mississippi, Oklahoma, Tennessee, Texas (decennial) : Forest acreage, timber volume, cut, growth, and mortality. Reporting unit: timber inventory plots. Unrestricted; 10 years; 200,000 cards.
- 40-A-48 Pulpwood production (annual) : Pulpwood procurement by county and type of wood-species, bolts, chips. Reporting unit: pulp mills. Confidential; 2 years; 4,000 cards.
- 40-A-49 Timber use other than pulpwood by State (decennial) : Timber volume removed from forests by species. Reporting unit: individual firm. Confidential; 10 years; 20,000 cards.
- 40-A-50 Stumpage Prices (selected intervals) : Characteristics of the stumpage offered and the price received. Reporting unit: individual sale. Unrestricted; permanent; 5,000 cards.

- 40-A-51 Small Private Forest Land Ownership Survey, Ohio and Missouri (one-time survey) : Study was to determine the ownership characteristics of the region, attitudes of owners toward their forest land, and if possible, reasons why owners adopt or do not adopt forestry practices and programs. Reporting unit: woodlot owners within four-county area in Ohio and Missouri. Confidential; permanent; 1958; 500 cards.
- 40-A-52 Input and Output Study of Hardwood Log Production (one-time survey) : Information on number, quality and costs of input and outputs for individual sawlogs. Reporting unit: individual boards. Confidential; 5 years; 1962; 20,000 cards.
- 40-A-53 Forage Production and Composition Survey (one-time survey) : Herbaceous and browse production and composition; overstory timber stand size, stocking class, site class, aspect position on slope, percent slope; livestock use; fire history; logging or TSI; erosion hazard index; current erosion index, soil stability rating; ground cover basal area and percent crown cover. Unrestricted; permanent; 1961; 40,000 cards.
- 40-A-54 The Market Potential for Residential Fencing, St. Louis (one-time survey) : Residential fence and home characteristics including: type of fence, volume of wood fence material, location of fence, age and type of home. Unrestricted; permanent; 1964; 600 cards.
- 40-A-55 Forest Inventory, Central States (decennial) : Volume and area information by state, county, and plot by species, dbh, ownership forest type, stand-size, site, tree class, stocking, grazing intensity, mortality, growth and quality. Reporting unit: sample plots and individual trees. Unrestricted; permanent; 150,000 cards.
- 40-A-56 Small Woodland Owners, Ohio (one-time survey) : Survey of factors such as participation in ACP woodland practices and participation in voluntary woodland practices. Reporting unit: woodland owners. Confidential; permanent; 1963; 520 cards.
- 40-A-57 Survey of Outdoor Recreation in Ohio (one-time survey) : Survey of owner's background characteristics, owner's conception of commercial outdoor recreation and owner's economic resources. Reporting unit: owners of forest recreation enterprises, such as picnic areas, camping areas, swimming beaches, riding stables, and pay lakes. Confidential; permanent; 1963; 280 cards.
- 40-A-58 Survey of Picnic Enterprises in Ohio (one-time survey) : Survey of factors such as location, size, natural or manmade attractions, services and facilities, costs of developing and operating picnic enterprises, income conversion surplus and conversion surplus ratio. Reporting unit: Forest Picnic Enterprises in Ohio. Confidential; 10 years; 1963; 72 cards.
- 40-A-59 National Compilation of Forest Survey Statistics (decennial) : By State; forest areas, inventories, annual growth, cut, and mortality. Reporting unit: Forest survey units at forest experiment stations. Confidential; permanent; 1962-63; 50,000 cards.
- 40-A-60 Wood Used by Manufacturers (selected intervals) : Amounts and species of wood used in manufacturing by industry, product, and forms of wood (lumber, bolts, veneer, plywood, hardboard, and particleboard). Reporting unit: sample survey of all manufacturing plants. Confidential; permanent; 1960; 84,000 cards.
- 40-A-61 Wood Preservation Statistics (annual) : Volume and species of wood products treated by different preservatives and fire retardants, and volume of chemicals used. Reporting unit: canvass of all wood-treating plants in the United States. Confidential; 5 years; 1959-63; 18,000 cards.
- 40-A-62 Wood Used in FHA Housing (selected intervals) : Amounts and forms of wood used by house part in FHA inspected houses. Reporting unit: sample survey of house types from a sample of FHA offices. Unrestricted; permanent; 1959 and 1962; 65,000 cards.
- 40-A-63 Working Circle Timber Inventory, Region 3 (decennial) : Data on individual sample trees, species, dbh, height, and class. Data on soil erosion, diseases, timber type, and logging information. Reporting unit: national forests. Unrestricted; 10 years; 1962; 80,000 cards; 400 tapes.

- 40-A-64 Timber Sales Inventory, Region 3 (selected intervals): Data on individual sample trees, species, dbh, height and class. Data on soil, erosion, diseases, timber type and logging information. Reporting unit: individual sale. Unrestricted; 6 to 12 years; 60,000 cards; 300 tapes.
- 40-A-65 Fire Weather Records, Intermountain (daily): Weather and fuel conditions. Reporting unit: individual station. Unrestricted; permanent; 1954 to present; 150,000 cards.
- 40-A-66 Lightning Research, Montana (selected intervals): Atmospheric electric field-lightning electrostatic field. Reporting unit: individual lightning discharges. Unrestricted; 5 years; 1960-61; 1,600 cards.
- 40-A-67 White Pine Study, Intermountain Region (one-time survey): An accounting of man-hours, vehicle miles, and costs for various phases of timber management activities. Reporting unit: individual timber management project. Confidential; 4 years; 1962-65; 2,500 cards.
- 40-A-68 Forest Survey of Northern Idaho (one-time survey): Covers area, volume, growth and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1960-64; 30,000 cards.
- 40-A-69 White Pine Study, Intermountain Region (one-time survey): An inventory of lands capable of growing western white pine, including a description of site quality, age and species of stands presently growing on these lands, management silvicultural data. Reporting unit: sample acres. Unrestricted; 7 to 10 years; 1961-63; 1,000 cards.
- 40-A-70 Forest Survey of Western Montana (selected intervals): Covers area, volume, growth, and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1953-58; 30,000 cards.
- 40-A-71 Forest Survey of Wyoming (selected intervals): Covers area, volume, growth, and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1957-60; 25,000 cards.
- 40-A-72 Forest Survey of Colorado (selected intervals): First survey of forests of Colorado. Covers area, volume, growth, and mortality of trees. Reporting unit: individual trees. Unrestricted; 7 to 10 years; 1956-59; 70,000 cards.
- 40-A-73 Forest Survey, Southeast Region (decennial): Area, volume, growth, mortality, and timber cut statistics. Reporting unit: individual acres and individual trees. Unrestricted; permanent; 2 million cards.
- 40-A-74 White Pine Log Grade Study, Northeast Region (one-time survey): Classification and measurement of lumber cut from individual logs; i.e., width, length, and grade. Reporting unit: individual boards. Unrestricted; 1956-61; 37,000 cards.
- 40-A-75 Economics of Ponderosa Pine Dwarfmistletoe Control (one-time survey): Rates of return on dwarfmistletoe control investment. Reporting unit: timber stand. Unrestricted; 1 year; 10,000 cards.
- 40-A-76 Diameter Distributions for Douglas-Fir Stands (one-time survey): Number of trees by diameter class. Reporting unit: individual tree stand. Unrestricted; 1 year; 500 cards.
- 40-A-77 Financial Yields from Hardwood Stand Conversion (one-time survey): Present worths associated with conversion of stand from alder to Douglas-fir. Reporting unit: individual timber stands. Unrestricted; 1 year; one tape.

## COMMODITY EXCHANGE AUTHORITY

- 40-B-1 Position Surveys—Specific Commodity Market (selected intervals): Futures holdings of each trader in the commodity market in which survey is conducted. Reporting unit: futures commission merchants, members of contract markets, foreign brokers. Confidential; 3 years; 14,000 cards.
- 40-B-2 Cash Commodity Positions (weekly): Cash positions of respondents holding futures positions in specific commodities. Reporting unit: merchandisers, processors, or dealers in grains, cotton, eggs. Confidential; 5 years; 64,000 cards.

- 40-B-3 Futures Positions of "Special Accounts" (daily): Identification of and classification of positions of large traders. Reporting unit: futures traders whose open positions have reached reportable size. Confidential; 5 years; 210,000 cards.
- 40-B-4 Futures Positions of "Special Accounts" (daily): Identification of special accounts and their reportable futures positions. Reporting unit: futures commission merchants and foreign brokers. Confidential.
- 40-B-5 Futures Trading and Open Contracts (daily): Futures transactions, and open contracts carried on books of exchange clearing members. Reporting unit: clearing members of contract markets. Confidential; 2 months; 54,000 cards.

## STATISTICAL REPORTING SERVICE

- 40-C-1 Wool Report (monthly): Pounds, proceeds, and head shorn by date of shearing and sale. Reporting unit: farmers applications to ASCS for wool incentive payments. Confidential; 3 years; 1962-64; 800,000 cards; 12 tapes.
- 40-C-2 Consumer Expenditure Survey (one-time survey): Income and expenditures for a period of 1 year. Reporting unit: rural households. Confidential; 10 years; 1961; 800,000 cards; 30 tapes.
- 40-C-3 Objective Yield Surveys (monthly): Acreage along with objective counts and measurements of plants and fruits. Reporting unit: a sample of farmers. Confidential; 1 year; 159 to present; 400,000 cards; 5 tapes.
- 40-C-4 Cold Storage Report (monthly): Storage capacity of warehouses and weights of over 80 commodities in storage. Reporting unit: all types of refrigerated warehouses. Confidential; 3 years; 1961 to present; 576,000 cards; 6 tapes.
- 40-C-5 Prices Paid Surveys—Feed (monthly): Number of reports, average price and tax, and estimated State price for 50 feeds. Reporting unit: State summary data compiled by SRS field offices. Confidential; permanent; 1958 to present; 144,000 cards; 10 tapes.
- 40-C-6 Prices Paid Surveys—Food and Clothing (monthly): Reports on farmer purchases of 80 to 100 food and clothing items. Reporting unit: individual reports for chainstores, State summary reports on independent clothing stores. Confidential; permanent; 1959-63; 160,000 cards; 24 tapes.
- 40-C-7 Slaughterhouse Survey (monthly): Number of head slaughtered, live and dressed weight, cost, and class of cattle, hogs, and sheep slaughtered. Reporting unit: a sample of federally inspected meatpacking plants. Confidential; 6 years; 1962 to present; 50,000 cards.
- 40-C-8 Nonfederally Inspected Slaughter (monthly): Number of head, average and total live weight by size groups. Totals combined with federally inspected slaughter. Reporting unit: State summary data compiled by SRS field officers. Confidential; 2 years; 1963-64; 12,000 cards.
- 40-C-9 Horticultural Specialties Survey (annual): Grower data on plants on hand, or in production, quantities sold, and value. Reporting unit: individual commercial growers in selected States. Confidential; 2 years; 1963-64; 24,000 cards; 2 tapes.
- 40-C-10 Prices Received by Farmers—Potatoes and Citrus (monthly): Revised monthly estimated prices and weights. Reporting unit: State estimates prepared by field offices and Crop Reporting Board. Confidential; 2 years; 1962-63; 8,000 cards.
- 40-C-11 Telephone and Electricity Survey (annual): Kilowatt-hours and electric bill, telephone bill, LP gas purchased and amount of bill. Reporting unit: a sample of farmers and prices they pay. Confidential; 6 years; 1961 to present; 114,000 cards; 6 tapes.
- 40-C-12 June-December Enumerative Survey (annual): Includes acreage of most crops, livestock by specie and class, farm numbers, etc. Reporting unit: farmers—a probability sample. Confidential; 5 years; 1961 to present; 1 million cards; 30 tapes.

- 40-C-13 Market Records (monthly): Receipts of cattle, hogs, sheep, and lambs by State of origin. Reporting unit: individual livestock market and packing plants. Confidential; 2 years; 1963-64; 20,000 cards; 2 tapes.
- 40-C-14 Beef Steer and Heifer Report (monthly): Number of head, weight, and cost by grades for steers and heifers. Reporting unit: each of the 14 major livestock markets. Confidential; 6 years; 1962 to present; 7,500 cards; 2 tapes.
- 40-C-15 Dairy Manufacturers Survey (monthly): Production of butter, ice cream and other frozen dairy products, cheeses, canned milk, etc. Reporting unit: plants manufacturing dairy products (16 States). Confidential; 2 years; 1962-63; 10,000 cards.
- 40-C-16 Building Values (one-time survey): Value of property of farm, residential on farm, other buildings. Reporting unit: individual farms. Confidential; 3 years; 1963; 4,000 cards.

## AGRICULTURAL STABILIZATION AND CONSERVATION SERVICE

- 40-D-1 Defense Data Program (selected intervals): Reporting unit: individual plant locations. Confidential; permanent; 40,000 cards.
- 40-D-2 Storage Résumé (monthly): Warehouse facilities by State showing location, commodity code, number of lots, units, and quantity in store by capital and appropriated fund. Unrestricted; 2 years; 18,000 cards.
- 40-D-3 Area Recap, Instore Processed Commodity Inventory (semimonthly): Area recap—State total by commodity, bulk, and/or packaged with overall total. IM3R commodity class by warehouse and location giving quantity by commodity and program year. Unrestricted; 2 years; 18,000 cards.
- 40-D-4 Position 4—Instore—Merchandising Inventory (semimonthly): Inventory listings showing lot number, commodity code, program code, program year, quantity, and warehouse in which stored. Reporting unit: inventory lot. Unrestricted; 2 years; 15,000 cards.
- 40-D-5 Position 3, Intransit, Positions 5, Ordered Not Shipped From Storage Inventory (semi-monthly): Position 3, reference number; commodity code, program code, lot number, quantity; position 5, reference number, commodity code, program code, program year, lot number, quantity. Unrestricted; 2 years; 50,000 cards.
- 40-D-6 Approved, Accepted and Reserve Warehouses (quarterly): Name and address; mailing address; plant address; type of storage, dry, cold, tank, or whey; number of cars loaded or unloaded per day; delivering carrier. Reporting unit: warehouses. Unrestricted; 2 years; 3,500 cards.
- 40-D-7 Current Warehouse Grain Activity (daily): Grain acquisitions, dispositions, and adjustments. Reporting unit: country and terminal warehousemen. Unrestricted; 30 days; seven tapes.
- 40-D-8 Grain Prices and Discount Formulas (daily): Base prices for barley, corn, flax, grain sorghums, oats, rye, wheat, and market premiums and discounts. Reporting unit: grain trade. Unrestricted; one tape.
- 40-D-9 Grain Inventory Open File (daily): Grains—barley, corn, flax, grain sorghums, oats, rye, wheat, and edible beans. Reporting unit: warehouse receipts. Unrestricted; 14 tapes.
- 40-D-10 Elevator Name and Address Master File (selected intervals): Elevator name, address, settlement markets, freight rates, storage, and loadout capacities. Reporting unit: warehouse. Unrestricted; 1961 to present; one tape.
- 40-D-11 CCC Commodity Loan Transactions (daily): Loans made, repayments, collateral, acquired, loans written off, and loans outstanding. Reporting unit: individual loan. Confidential; 60 days; 120 tapes.
- 40-D-12 Commercial Warehouses Under UGSA—21 States (daily): Warehouse name, address, and capacity. Reporting unit: warehouse. Confidential; permanent; 4,600 cards.
- 40-D-13 CCC-owned Grain Bin and Equipment Facilities (daily): Purchases, transfers between States, counties, and bin-site locations, and dispositions. Reporting unit: individual facility. Confidential; 97 days; three tapes.

- 40-D-14 CCC-owned Grain Inventory in CCC Bin Sites (daily): Grain receipts and withdrawals. Reporting unit: individual binsite. Confidential; 14 days; one tape.
- 40-D-15 CCC-Commodity Inventories (daily): Month-end inventory; acquisitions and dispositions during month. Reporting unit: internal vouchers supporting the general ledger. Confidential; 1961 to present; 100,000 cards.
- 40-D-16 CCC Warehouse—Stored Grain Dispositions, 21 States (daily): Commodity, trust member, quality and quantity of grain dispositions. Reporting unit: warehouse receipt. Confidential; 7 days; one tape.
- 40-D-17 CCC Warehouse—Stored Grain Acquisitions—21 States (daily): Commodity, quality, quantity of grain acquired. Reporting unit: warehouse receipts. Confidential; 7 days; four tapes.
- 40-D-18 CCC Warehouse—Stored Grain Inventories—21 States (daily): Description of CCC grain inventory stored in commercial warehouses. Reporting unit: warehouse receipt. Confidential; 14 days; 14 tapes.

## AGRICULTURAL MARKETING SERVICE

- 40-E-1 Tobacco Stocks Report (quarterly): Stocks of leaf tobacco owned by dealers and manufacturers. Reporting unit: tobacco dealers and manufacturers. Confidential; 5 years; 1960 to present; 50,000 cards.
- 40-E-2 Truck Shipments of Fresh Fruits and Vegetables for California and Florida (daily): Package units of fresh fruits and vegetables. Reporting unit: carlot inspections. Unrestricted; 90 days.
- 40-E-3 Unloads of Fresh Fruits and Vegetables in 41 Cities (daily): Carlots and carlot equivalents of fresh fruits and vegetables by commodities and origin (States or countries) unloaded in 41 principal market cities. Reporting unit: carlot inspections. Unrestricted; 30 days.
- 40-E-4 Rail Shipments of Fruits and Vegetables (monthly): Carlot units of fresh fruits and vegetables by waybilling stations, commodities, and origin. States or countries. Reporting unit: carlot inspections. Unrestricted; 1 year.
- 40-E-5 Fruit and Vegetable Rail Shipments (daily): Carlot units of fresh fruits and vegetables by commodities and origin (States or countries). Reporting unit: carlot inspections. Unrestricted; 90 days.
- 40-E-6 Egg Products (Liquid, Frozen, and Dried) Report (weekly): Quantities produced. Reporting unit: plants under Federal grading. Confidential; 5 years; 1962 to present; three tapes.
- 40-E-7 Poultry Canning Report (monthly): Quantities of poultry used in cutting up and further processed and quantities condemned. Reporting unit: further processing plants under Poultry Products Inspection Act. Confidential; 5 years; 1960 to present; 13,000 cards; two tapes.
- 40-E-8 Slaughter and Evisceration Report (weekly): Quantities of poultry inspected and condemned. Reporting unit: slaughter and evisceration plants under Poultry Products Inspection Act. Confidential; 5 years; 1959 to present; 100,000 cards; four tapes.
- 40-E-9 Milk Marketing Program 9110 (one-time survey): Volume weights of milk and milk products. Reporting unit: dairy plants. Confidential; permanent; 1961-62; two tapes.
- 40-E-10 Milk Marketing Product Reports (monthly): Sales of fluid milk products. Reporting unit: dairy plants. Confidential; permanent; 1960 to present; 16 tapes.
- 40-E-11 Milk Marketing Price Report—MO-1 (monthly): Milk receipts, utilization, and prices. Reporting unit: dairy plants. Confidential; permanent; 1960 to present; 40 tapes.
- 40-E-12 GR-132, Volume of Grain Inspections (annual): Volume of each kind of grain inspected at each market, by movement. Reporting unit: grain elevators. Unrestricted; 3 years; 75,000 cards.
- 40-E-13 Grain Quality (Formerly Grain Inspections) Data (annual): A systematic sample for estimating quality of each grain crop. Reporting unit: grain elevators. Unrestricted; 3 years; 300,000 cards.

- 40-E-14 Cotton Quality Survey (annual): Fiber and processing properties of model qualities of cotton produced in the United States. Reporting unit: cotton gins. Unrestricted; permanent; 1946 to present; 36,000 cards.

## ECONOMIC RESEARCH SERVICE

- 40-F-1 Social Security Survey (annual): Social security benefits and how employed by farmers. Reporting unit: farmers covered by social security. Confidential; 5 years; 149,000 cards.
- 40-F-2 Use of Highways in Crop Disposal (one-time survey): Values of varying-type highways in expediting various type crops. Reporting unit: highway commission, truckers, farmers, co-ops, markets. Confidential; 5 years; 1964; 24,000 cards.
- 40-F-3 Surveys of Agricultural Finance (selected intervals): Various lending operations, loan surveys, etc., for farmers. Reporting unit: lending institutions, Federal Reserve, individuals, etc. Confidential; 1958-61; 450,000 cards.
- 40-F-4 Great Plains Survey (one-time survey): Landownership, water rights, mineral rights—methods of obtaining ownership. Reporting unit: individual farmers. Confidential; 5 years; 1960; 32,000 cards.
- 40-F-5 Fairfax Tax Study (one-time survey): Comparative tax assessments on former farmer area, which is now tax-classified "suburban." Reporting unit: individual tax assessments. Confidential; 2 years; 2,000 cards.
- 40-F-6 Great Plains Pricing Survey (one-time survey): Factors considered by sellers and buyers in pricing farm real estate. Reporting unit: individual sellers and buyers. Confidential; 5 years; 1960; 20,000 cards.
- 40-F-7 Great Plains Survey (Farmers Living Standards) (one-time survey): Amount of land, living standards, size of family, etc. Reporting unit: individual farmers. Confidential; 5 years; 1960; 46,000 cards.
- 40-F-8 Transportation Study (one-time survey): Transportation costs of corn between demand and supply points. Reporting unit: transportation units of various agencies. Confidential; 5 years; 1963; 10,000 cards.
- 40-F-9 Economics of Housing for Migrant Hired Farmworkers (one-time survey): Economics of housing for migrant hired farmworkers, social security information of farmers and farmworkers. Reporting unit: farmers and farmworkers. Confidential; 3 years; 1963; 13,000 cards.
- 40-F-10 Farm Real Estate Values (one-time survey): Values of real estate. Reporting unit: individual farmers. Confidential; 5 years; 1960; 26,000 cards.
- 40-F-11 Tax Survey (annual): Farm real estate taxes. Reporting unit: State tax offices. Unrestricted; 5 years; 1960-63; 150,000 cards.
- 40-F-12 Farm Real Estate Market Survey (semiannual): supply of and demand for farmland, actual sales of farmland, current prices and expected price movements, type of buyers and sellers, availability. Reporting unit: individuals. Unrestricted; permanent; 1956-64.
- 40-F-13 Voluntary Home Mortgage (one-time survey): Insurance of home mortgage loans made by banks, home finance companies, leading agencies, insurance companies, etc. Reporting unit: individual loans. Confidential; 1 year; 1963; 115,000 cards; two tapes.
- 40-F-14 Building Value Survey (one-time survey): Individuals asked to estimate value: their entire farm, all buildings, residence. Reporting unit: individual farmers. Confidential; 3 years; 1963; 12,000 cards; two tapes.
- 40-F-15 Credit Survey—Dairy in Wisconsin; Hog-Beef in Corn Belt (one-time survey): Financial standing of individuals, including dairy operators in eastern Wisconsin and hog-beef feeders in the Corn Belt. Reporting unit: individuals. Confidential; 3 years; 1963; 2,000 cards.

- 40-F-16 Extent of Spraying and Dusting on Farms (one-time survey): Individual farm operations in connection with chemical treatment for insect, disease, and weed control. Reporting unit: individual farmers. Confidential; 5 years; 1958; 44,000 cards.
- 40-F-17 Machinery and Equipment Study (one-time survey): Survey of farm tractors and machinery. Reporting unit: farmers. Confidential; permanent; 1957; 10,000 cards.
- 40-F-18 Illinois Feeder Cattle Study (one-time survey): Feed conversion data for droves of cattle fed by Illinois farm cooperatives. Reporting unit: individuals. Confidential; permanent; 1938-63; 9,000 cards.
- 40-F-19 Methods Used to Distribute Fertilizer (one-time survey): Reporting unit: individual farmers. Confidential; 5 years; 1962; 48,000 cards; two tapes.
- 40-F-20 Harvesting the Hay Crop (one-time survey): Survey of farmers' operations in the harvesting of hay. Reporting unit: individual farmers. Confidential; 5 years; 1961; 90,000 cards.
- 40-F-21 Recordings of Farm Mortgages (biennial): Characteristics of farm mortgages recorded (closed): interest rates, term, size averages. Reporting unit: lenders. Confidential; 1 year; 1963; 56,000 cards; three tapes.
- 40-F-22 Liquid Petroleum Fuel Used by Farmers (one-time survey): Farmers and extent of their use of different types of fuel. Reporting unit: individual farmers. Confidential; 5 years; 1959; 64,000 cards.
- 40-F-23 Harvesting Small Grains and Field Shelling Corn (one-time survey): Extent of different harvesting methods used on small grains and extent of field shelling of corn. Reporting unit: individual farmers. Confidential; 5 years; 1960; 40,000 cards.
- 40-F-24 OECD Agricultural Exports and Imports (quarterly): SITC commodities exported and imported by country of destination and origin. Reporting unit: SITC commodities. Unrestricted; permanent; 1963; eight tapes.
- 40-F-25 State Export Equivalent Study (one-time survey): State share of production of U.S. exports and imports of selected commodities. Unrestricted; permanent; 1960-61; 10,000 cards.
- 40-F-26 SRS June Enumerative Survey, Farm Population Data (annual): Color of operator, number of persons in operator's household, number of other households on farm. Farm classification items, such as size of farm, value of products sold, etc. Reporting unit: farm operator households. Confidential; permanent; 1960-62; 18,000 cards; two tapes.
- 40-F-27 Maryland Suburbanization Study (one-time study): Characteristics of head of household and family members, information on changes in community undergoing rapid suburbanization, and attitudes of persons toward these changes. Reporting unit: households in urban fringe. Confidential; 5 years; 1960; 4,000 cards.
- 40-F-28 1/1000 Sample, Population, and Housing (one-time survey): Area and unit identification. Characteristics of persons, households, families, subfamilies, associated persons, mothers of children under 18, and housing units. Reporting unit: households and individuals within households. Unrestricted; permanent; 1960; seven tapes.
- 40-F-29 Textile Imports (monthly): Import statistics, giving commodity codes, total quantities imported, and value for cotton, wool, and the manmade fibers. Reporting unit: total imports for consumption by textile commodity reported by the Bureau of Census. Unrestricted; permanent; 1963-64; nine tapes.
- 40-F-30 Gross Farm Income (annual): Cash receipts from farm marketings; value of home consumption; value of annual change in farm inventories; index numbers of the volume of farm marketings and home consumption. Reporting unit: secondary data from governmental agencies. Unrestricted, permanent; 1949-63; 650,000 cards.
- 40-F-31 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey): Data card. Expanded carlots, firm size and market structure. Reporting unit: wholesale firms. Confidential; 4 years; 1958-59; 114,000 cards.
- 40-F-32 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey): Employees IRS forms 1065 and 1120. Reporting unit: IRS. Confidential; 4 years; 1959-60; 32,000 cards.

- 40-F-33 Rio Grande Tomato Study, ME 3-67 (one-time survey) : Data cards, structure of lower Rio Grande tomato market. Reporting unit: wholesale firm. Confidential; permanent; 1960-61.
- 40-F-34 Rio Grande Citrus Study, ME 3-67 (one-time survey) : Data card, structure of Rio Grande Citrus Market. Reporting unit: wholesale firms. Confidential; permanent; 1960-61; 75,000 cards; eight tapes.
- 40-F-35 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey) : Data cards, unexpanded carlots and firm size. Reporting unit: wholesale firms. Confidential; permanent; 1958-59; 21,000 cards.
- 40-F-36 Market for Food in Schools (one-time survey) : School characteristics, type of lunch program, and foods used. Reporting unit: public and private elementary and secondary schools. Confidential; 5 years; 1962-63; 10,000 cards.
- 40-F-37 Public Food Distribution Programs Research (selected intervals) : Food consumption and marketing information plus income and other characteristics of low income families. Reporting unit: low income households in: Baltimore, Md., Detroit, Mich., and urban and rural areas of Fayette County, Choctaw County, Okla., Escambia County, Fla., and St. Louis, Mo. Confidential; permanent; 1961-present; 150,000 cards.
- 40-F-38 Convenience Food Study (monthly) : Cost and time per serving of convenience foods and home-prepared foods. Reporting unit: supermarkets. Unrestricted; 3 years; 1959-60; 424 cards.
- 40-F-39 Low-Fat Milk Study (one-time survey) : Low-fat milk sales, composition of low-fat milk, prices for December 1962. Percent low-fat milk represents of total whole, low-fat and skim sales. Reporting unit: milk processors. Confidential; 3 years; 1962; 3,000 cards.
- 40-F-40 Marketing Horticultural and Special Crops, Promotional Practices (one-time survey) : Current marketing practices by retail florists (advertising, promotion, merchandising, pricing, etc.). Reporting unit: retail florists. Confidential; 3 years; 1964; 35,000 cards.
- 40-F-41 Dairy Promotion Study (monthly) : Pounds of milk sold per capita and prices and display size. Reporting unit: Federal Milk Order Markets and Retail Food Stores. Confidential; 3 years; 1963-65; 50,000 cards.
- 40-F-42 Expenditures for Promotion (one-time survey) : Sums spent for promotion and sources of revenue. Reporting unit: farm commodity groups. Confidential; 3 years; 1962-63; six tapes.
- 40-F-43 Food Stocks in Away-From-Home Eating Establishments (one-time survey) : Inventories of food and beverage products. Reporting unit: establishments that generally serve food for onpremise consumption. Confidential; 3 years; 1964; five tapes.
- 40-F-44 Flexibility of Dairy Manufacturing plants (annual) : Production of products by plants. Reporting unit: all plants manufacturing dairy products in United States. Confidential; 4 years; 1961.
- 40-F-45 Name and Address of Egg Assemblers (one-time survey) : Name and address of egg assemblers. Reporting unit: egg assembler and brokers. Unrestricted; 1 year; 1957-58; 26,000 cards.
- 40-F-46 McClain Cost Data for Dairy Plants (quarterly) : Sales and costs by items for 70 fluid milk plants. Reporting unit: individual plants. Confidential; permanent; 1959-63; 500,000 cards.
- 40-F-47 Egg Quality Study (one-time survey) : Gradeouts, yields of producers and price received for eggs. Reporting unit: egg packing plants. Confidential; 2 years; 1960-61; 10 tapes.
- 40-F-48 Weighted Meat Prices (monthly) : Retail prices of beef, veal, pork, and lamb. Reporting unit: cooperating chainstore. Confidential; 3 years; 1962-64; 4,000 cards.
- 40-F-49 Financial Statistics of Food Manufacturers (annual) : Advertising expenditures, total costs, total sales, total net income from IRS Source Book. Reporting unit: asset size classes. Unrestricted; permanent; 1953-60; 4,000 cards.
- 40-F-50 Interindustry Input-Output Data (one-time survey) : Outputs by producing industry and by consuming industry in producer prices. Reporting unit: industry. Unrestricted; permanent; 1947; 40,000 cards.

- 40-F-51 Farm Value, Retail Price, and Farm-Retail Spread for Food Products (monthly) : Farm value, retail price, and farm-retail spread. Reporting unit: national average by commodity. Unrestricted; permanent; 1947-58; 2,000 cards.
- 40-F-52 Wool Classification Study (one-time survey) : Wool quality factors and prices. Reporting unit: wool warehouse. Unrestricted; 6 years; 1957-60; 60,000 cards.
- 40-F-53 Central Market Study (cotton) (weekly) : Date of sale, location, volumes, price, quality. Reporting unit: individual respondents. Unrestricted; 6 years; 1959-61; 288,000 cards.
- 40-F-54 Study of Cotton Warehouses Storage Cost (one-time survey) : Cotton storage costs. Reporting unit: warehouses. Confidential; permanent; 1959-60; 2,000 cards.
- 40-F-55 Grain Storage and Handling Costs (one-time survey) : Operating costs for grain elevators. Reporting unit: grain elevator operators and owners. Confidential; permanent; 1959-61; 8,000 cards.
- 40-F-56 Wool Market News Study (weekly) : Wool prices and quality factors, location. Reporting unit: individual sales. Unrestricted; 6 years; 1962-63; 530 cards.
- 40-F-57 Census Data (quinquennial) : Acres and production for 70 crops and 10 livestock items. Reporting unit: farmer respondents, from published census data. Unrestricted; permanent; 1959; 300,000 cards.
- 40-F-58 Normalizing Study (biennial) : Acres planted, harvested, yield, production, price and value for 70 crop and 10 livestock items. Reporting unit: States. Unrestricted; permanent; 1939-62; 52 tapes.
- 40-F-59 Ohio River Basin Study (one-time survey) : Estimated yields (two levels) for land capability units within land resource areas. Reporting unit: Work unit conservationists, land capability units within land resource area. Unrestricted; permanent; 1963; 33 tapes.
- 40-F-60 Conservation Needs Inventory Land Capability and Use Data (one-time survey) : Land use in 1958 and estimated land use in 1975 by land capability subclass. Reporting unit: county committees, counties. Unrestricted; permanent; 1958; 7,500 cards.
- 40-F-61 Watershed Project Needs (CNI) (one-time survey) : Acreages requiring project action for flood control, erosion control, drainage and irrigation. Reporting unit: county committees, watersheds less than 250,000 acres. Unrestricted; permanent; 1958; 2,000 cards.
- 40-F-62 Conservation Needs, 160-Acre Sample Plots (one-time survey) : Land use, soil type, soil slope, antecedent erosion, and capability class for Oklahoma, Louisiana, New Mexico, and Arkansas. Reporting unit: 160-acre sample plots (2 percent of total area). Unrestricted; permanent; 1958; 60,000 cards.
- 40-F-63 Ownership of Farm Land in the United States (one-time survey) : Characteristics of owners, acreage owned, method of acquisition and disposition. Reporting unit: landowners, random sample each county in each State. Unrestricted; permanent; 1946; 38,000 cards.
- 40-F-64 Relative Efficiency of Alternative Tenure Arrangements (one-time survey) : Tenure, and farm input and output data. Reporting unit: farm operators, sample in Nebraska, Kansas, Iowa, and Missouri. Unrestricted; 10 years; 1953-57; 16,000 cards.
- 40-F-65 Land Ownership in the Southeastern States (one-time survey) : Personal characteristics of landowners, amount and use of land owned, and changes in land used from 1955 to 1960. Reporting unit: Landowners, sample segments within counties of seven States. Unrestricted; permanent; 1955-60; 18,000 cards.
- 40-F-66 Land Ownership in the Great Plains States (one-time survey) : Personal characteristics of landowners, amount and use of land owned. Reporting unit: landowners, sample counties of 10 Great Plains States. Unrestricted; permanent; 1957; 153,000 cards.
- 40-F-67 Study of Feed Grain Program (one-time survey) : Cropland use, livestock programs, and factors relating to 1961 feed grain program. Reporting unit: farms in Iowa. Unrestricted; 5 years; 1961; 4,000 cards.



- 40-F-68 Appraisal of Soil Bank Programs in Selected Areas of Georgia (one-time survey): Characteristics of participants and nonparticipants in soil bank program. Reporting unit: farm operators in six sample counties in Georgia (both participants and nonparticipants in the soil bank program). Confidential; 10 years; 1956-59; 12,000 cards.
- 40-F-69 Resource Use and Incomes of Farm Families in Georgia (one-time survey): Family characteristics (age, education, occupation, income) and farm characteristics (land use, production, livestock numbers, nonfarm work, farm income). Reporting unit: dwelling units in the open country or rural households. Confidential; 10 years; 1957; 12,000 cards.
- 40-F-70 Insect Control and Related Cotton Practices Study (one-time survey): Information on whether or not cotton insect control practices were used, and if so, the acres covered, the kind, quantity, type, and rate of application, the cost of insecticide used and method of application. Similar information on pre-emergence and post-emergence herbicides, defoliant and fertilizer, and also estimates of the 1961 and 5 year average yields of cotton. Reporting unit: individual farmers. Confidential; 10 years; 1961; 16,500 cards.
- 40-F-71 An Inventory of Land and Soil Resources in Pennsylvania (one-time survey): Acres of land by use, soil type, slope, and degree of erosion. Two percent random sample. Unrestricted; permanent; 1957-58; one tape.
- 40-F-72 Assessment and Taxation of Farmland—Rochester, N.Y. (one-time survey): Property and owner characteristics. Reporting unit: farmers in towns of Brighton, Henrietta, Rush, Avon, and Genesee. Confidential; permanent; 1963; 440 cards.
- 40-F-73 Market Egg Poultry Farm Adjustments (one-time survey): Labor estimates for poultry farm operations, egg production, replacement data, buildings, and equipment requirements, costs and returns. Reporting unit: market egg producers in Connecticut. Confidential; 3 years; 1960-61; 2,000 cards.
- 40-F-74 Yield Data (selected intervals): Acreage planted, acreage harvested and production for each crop for each county for each year. Reporting unit: all North Dakota counties. Unrestricted; permanent; 25,000 cards.
- 40-F-75 Southwest North Dakota Regrassing Study (one-time study): Description of farm, machinery inventory, livestock inventory, many other items as reported on the survey schedule. Reporting unit: farms. Confidential; 3 years; 1957-63; 3,000 cards.
- 40-F-76 ASCS Sample of Farms (one-time survey): Acres of cropland, non-cropland, wheat allotment, feed-grain base and normal conserving base on a 10-percent sample of farms in 32 counties in North Dakota. Reporting unit: farms. Unrestricted; 5 years; 1959-60; 3,600 cards.
- 40-F-77 Montana State-Lease Yield Data (biennial): State-lease number, location code, seeded acres, total production, and yields for spring, winter wheat, and/or barley. Reporting unit: unpublished yearly yield data obtained from the Montana Agricultural State-lease Records. Unrestricted; permanent; 1938-62; 10,000 cards.
- 40-F-78 Yearly County Yield Data, Montana (biennial): County code, crop code, year planted acres, harvested acres, production, yield per planted acre, yield per harvested acre. Reporting unit: yearly yield data obtained from the Montana Agricultural Statistics. Unrestricted; permanent; 1919-61; 20,000 cards.
- 40-F-79 Platte Valley Farmer-Livestock Feeder Survey (selected intervals): Location within county, number of cattle and sheep fed, acres of irrigated cropland, acres of sugar beets. Reporting unit: farmers and livestock feeders. Unrestricted; 6 years; 1953, 1959; 9,000 cards.
- 40-F-80 Farm Adjustments on Wheat Farms (one-time survey): Costs to produce wheat, barley, corn, sorghum, and grazing livestock for 1960 and estimated for 1970. Reporting unit: farm enterprise cost data. Confidential; 5 years; 1960, 1970; 500 cards.

- 40-F-81 Farm Machinery Costs by Size of Farm (one-time survey): Implement, ownership costs, repair costs, use and estimated life of farm machines. Reporting unit: individual farmers on various sizes of farms. Confidential; 1960; 6,000 cards.
- 40-F-82 Great Plains Survey (one-time survey): Farm size, land tenure, land values, land use, and inventories. Reporting unit: individual farm. Unrestricted; 1957; 12,000 cards.
- 40-F-83 ASCS Survey, South Dakota (one-time survey): Farm size, land tenure, limited land use, livestock, and machinery inventory. Reporting unit: individual farm. Unrestricted; 1952; 20,000 cards.
- 40-F-84 Representative Farms in Indiana (one-time survey): Land, livestock, machinery, buildings, capital, farming practices and plans. Reporting unit: individual farm. Confidential; 5 years; 1962; 12,000 cards.
- 40-F-85 Wisconsin Farmers Home Administration Data (annual): Farm size, production, income and cost data. Reporting unit: farm and home records-individual borrowers. Confidential; permanent; 1957-62; 12,000 cards.
- 40-F-86 The Michigan Farm Credit Panel (one-time survey): Physical farm data and farm and family financial data. Reporting unit: farm families. Confidential; 2 years; 1961 3,000 cards.
- 40-F-87 NC-54 Feed Grain-Livestock Study (one-time survey): Resources, personal characteristics, and financial position. Reporting unit: Stratified random sample of farms in lower Michigan and Northern Indiana. Confidential; permanent; 1960; 3,600 cards.
- 40-F-88 Lake States Dairy Study (one-time survey): Farm resources and enterprises, personal characteristics of operator. Reporting unit: random sample farmers in lower Michigan. Confidential; permanent; 1958; 3,600 cards.
- 40-F-89 Lake States Dairy Study—Michigan Drainage Study (one-time survey): Resurvey of 1959 dairy study farms in 2 areas, resources enterprises, drainage conditions, inventory, personal characteristics of operator. Reporting unit: individual farm. Confidential; permanent; 1962; 2,500 cards.
- 40-F-90 Census Data Hay and Silage Yields (selected intervals): Total tons and total acres of each hay and silage crop by counties in the United States from U.S. Census of Agriculture for years 1949 and 1954. Reporting unit: total tons and acres of each hay and silage crop in U.S. census by counties for 1949 and 1954. Unrestricted; 4 years; 1949-54; 20,000 cards.
- 40-F-91 North Central Iowa Farm Building Survey (one-time survey): Inventory of buildings and their use, cost, repairs, crop and livestock production. Reporting unit: farmers in north-central Iowa. Confidential; 15 years; 1963; 25,000 cards.
- 40-F-92 Feed Grain Program Study (one-time survey): Cropland use, livestock, practices in farming, factor relating to participation in feed grain program. Reporting unit: farms in Pacific Northwest and Corn Belt and Texas. Unrestricted; 5 years; 1963; 30,000 cards.
- 40-F-93 Survey of Pilot Cropland Conversion Program in Iowa (one-time survey): Cropland use in 1962 and 1963, livestock programs, reasons for and effect of participation in 1963 cropland conversion program. Reporting unit: farms in Polk and Dallas Counties, Iowa. Unrestricted; 5 years; 1963; 1,500 cards.
- 40-F-94 Economic Evaluation of Alternative Systems on Corn Belt Farms (one-time survey): Costs in harvesting, hauling, drying conditioning storage, and utilization of shelled corn. General farm organization, detailed account of corn production, record of labor and equipment. Reporting unit: Illinois farmers. Confidential; 5 years; 1961-62.
- 40-F-95 An Economic Appraisal of the Use of Water for Irrigation on Illinois Farms: Confidential; 5 years.
- 40-F-96 Minimum Resources for Specified Incomes (one-time survey): Costs and returns detailed in such a manner as to permit selection of minimum resources to produce incomes of \$2,500, \$3,500, \$45,000, and \$5,500 on corn farms and hog farms with land prices established at three different levels. Reporting unit: budgeted data for corn farms and hog farms of specific income levels. Unrestricted; 5 years; 1959-63.

- 40-F-97 Field Shelling Mechanical Drying and Storing Shelled Corn in Illinois (one-time survey): Mail questionnaire on punchcards. General characteristics of farms with specific information on corn harvesting, drying, and storage equipment. Reporting unit: data from 301 farms. Unrestricted; permanent; 4,000 cards.
- 40-F-98 Crop Yields, Acreage, Pires, and Gross Income (annual): Data by counties in Illinois. Reporting unit: Illinois Cooperative Crop Reporting Service. Unrestricted; permanent; 1925-63; 6,000 cards.
- 40-F-99 Livestock and Poultry Numbers (annual): Data by counties in Illinois. Reporting unit: Illinois Cooperative Crop Reporting Service. Unrestricted; permanent; 1925-63; 6,000 cards.
- 40-F-100 Feed Grain Program in Western Ohio (one-time survey): Crop acreages, yields, animal-unit of livestock, certain crop costs, fertilizer use, etc. Reporting unit: 160 sample farms. Unrestricted; 3 years; 1962; 2,200 cards.
- 40-F-101 Farm Size and Cotton Allotment Data—Missouri Delta (one-time survey): Acres of farmland, acres of cropland, and acres of cotton allotment. Reporting unit: Agricultural Stabilization and Conservation Service farm contract units. Confidential; 5 years; 1963; 10,000 cards.
- 40-F-102 The Effect of Selected Weather Variables on Corn Yields (decennial): Corn yields, date of planting, date of tasseling, drought-day, precipitation, and average temperature. Reporting unit: Columbia and Sikeston, Mo. Unrestricted; 5 years; 1955-63; 3,000 cards.
- 40-F-103 Data From Soils Testing (one-time survey): Physical characteristic of soil, crop yields, and fertilizer treatment. Reporting unit: information sheet for soil samples reported by farmers for each soil sample. Confidential; 8-10 years; 1956-60; 4,500 cards.
- 40-F-104 Conservation Needs Inventory Data for Arkansas (one-time survey): Principal physical characteristics of the land, such as slope, soil capability, soil type, and land use and the acres associated with each characteristic. Reporting unit: randomly selected plots 40 or 160 acres in size. Unrestricted; 8-10 years; 1958; 30,000 cards.
- 40-F-105 Capital Requirements and Ownership Costs, Arkansas Rice Farms (selected intervals): Location, tenure, size unit, enterprise organization, machine inventory, replacement practices, etc. Reporting unit: individual farm operator. Confidential; 8-10 years; 1959, 1961; 1,000 cards.
- 40-F-106 Organization and Operation of Texas Rice Farms (one-time survey): Location, tenure, size unit, major land use, enterprise organizations, requirements, and returns. Reporting unit: individual farm operators. Unrestricted; 8-10 years; 1960; 1,000 cards.
- 40-F-107 Input-Output Data, Texas, Crop and Livestock Farm Enterprises: Data developed pertains to farm resource restrictions, crop and livestock production requirements, yield levels, and product prices. Reporting unit: no respondents, data worked up from publications pertaining to resource requirements, production costs, and yields of crop and livestock farm enterprises. Unrestricted; 1 year; 1957.
- 40-F-108 Oregon Wheat Study (one-time survey): Total acreage, wheat acreage, allotment, and normal yield from wheat listing sheets. Reporting unit: individual farms. Unrestricted; permanent; 1953-59; 400 cards.
- 40-F-109 ASCS Data on Farms in Southeastern Idaho (one-time survey): Total land, cropland, wheat allotment, feed grain base, etc. Reporting unit: farm firms. Unrestricted; 3 years; 1963; 2,500 cards.
- 40-F-110 Sample Survey, Farms in Southeastern Idaho (one-time survey): Land by types, crops grown, livestock, and livestock facilities. Reporting unit: farm firm. Unrestricted; 3 years; 1963; 1,200 cards.
- 40-F-111 The Farm Work Force in Kern County, Calif. (one-time survey): Characteristics of workers, employment, earnings, seasonality. Reporting unit: Farm Production Economics Division, ERS, USDA, University of California, Davis, Calif. Unrestricted; 1961; 3,800 cards.

- 40-F-112 The Farm Work Force in Stanislaus County, Calif. (one-time survey): Characteristics of workers, employment, earnings, seasonality, migration. Reporting unit: Farm Production Economics Division, ERS, USDA, University of California, Davis, Calif. Unrestricted; 1962-63; 5,100 cards.
- 40-F-113 Irrigation Characteristics, Salinas Valley, Calif. (one-time survey): Depth, pump lift, horsepower, discharge of each well. Reporting unit: individual well. Confidential; permanent; 1950-63; 3,000 cards.
- 40-F-114 Oregon Cattle Price Data (monthly): Auction market location, type of cattle, average weight, price paid, date. Reporting unit: livestock auction markets. Confidential; 1964 to present; 2,500 cards.
- 40-F-115 Average Daily Gain and Feed Consumption, Cattle in Arizona (one-time survey): Average daily gain, daily feed consumption, age, weight, and breeds of cattle, types of feed fed, etc. Reporting unit: farmers (Arizona cattle feeders). Confidential; 1 year; 1960-61; 48 cards.

## DEPARTMENT OF COMMERCE

## BUREAU OF CENSUS

- 41-A-1 National Location Code File (selected intervals): Card file and tape file: place name, population total, geographic coordinates, census tract codes. Card file only: place name, population total, geographic coordinates, Universal Transverse Mercator (UTM) grid, reporting unit: standard location areas—tract and pseudo-tract. Unrestricted; permanent; 1960; 152,000 cards; two tapes.
- 41-A-2 Population Concentration (one-time survey): Place name, geographic coordinates, population of selected points, total population within 50 miles. Reporting unit: selected cities in the United States with a total population of 25,000 or more. Unrestricted; permanent; 1960; two tapes.
- 41-A-3 County City Data Books (selected intervals): Area, population, housing, vital statistics, manufacturers, trade, agriculture. Reporting unit: county, SMSA, urbanized area, unincorporated urban place, city of 25,000 or more. Unrestricted; permanent; 300,000 cards; three tapes.
- 41-A-4 Census of Agriculture County Summary (quinquennial): County totals for a variety of inventory, production, expenditure, and sales items—also data about farm operator and farm. Reporting unit: farms within county; 100 and 20 percent tabulations. Confidential; 6 years; 1959; 44,000 cards.
- 41-A-5 Census of Agriculture/Sample and Specified Farm Cards (quinquennial): Data on farms, farm characteristics, livestock and products, crops, fruits, values, etc. Reporting unit: farms. Confidential; 6 years; 1959; 8,900,000 cards.
- 41-A-6 Census of Agriculture (quinquennial): Data on farms, farm characteristics, livestock and products, crops, fruits, values, etc. Reporting unit: farms. Confidential; permanent; 1964.
- 41-A-7 Special Census of Metropolitan Louisville (one-time survey): Postal addresses and geographic and housing control items. Reporting unit: persons in households. Unrestricted; permanent; 1964; 300 tapes.
- 41-A-8 Census of Population and Housing (decennial): Major characteristics of population and housing. Reporting unit: persons in households. Confidential; permanent; 1960; 732 tapes.
- 41-A-9 Census of Housing—25 percent sample (decennial): Characteristics of occupied and vacant housing units. Reporting unit: housing unit. Confidential; permanent; 1960; 1,474 tapes.
- 41-A-10 Census of Population—Tallies—25 and 5 Percent Samples: Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 4,072 tapes.
- 41-A-11 Census of Population—Basic Records for the 25 and 5 Percent Samples: Social and economic characteristics of persons, families, and households. Reporting unit: persons. Confidential; permanent; 1959-60; 7,297 tapes.

- 41-A-12 Census of Population—One in a Thousand and One in Ten Thousand Samples (decennial): Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 13 tapes.
- 41-A-13 Census of Population—One in a Thousand and One in Ten Thousand Samples (decennial): Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 198,000 cards; 7 tapes.
- 41-A-14 Censuses—Control and Identification Tapes (decennial): Names of geographic entities for controlling and processing the censuses. Reporting unit: minor civil divisions. Unrestricted; permanent; 418 tapes.
- 41-A-15 Survey of Residential Alterations and Repairs (quarterly): Types and costs of residential alterations and repairs. Reporting unit: housing units. Confidential; permanent; 1959-63; 360,000 cards; 10 tapes.
- 41-A-16 Housing Vacancy Survey (monthly): Housing characteristics and vacancy status of vacant units. Reporting unit: vacant housing units. Confidential; permanent; 1959 to present; 100,000 cards; 120 tapes.
- 41-A-17 Congressional Districts (selected intervals): Social and economic characteristics and housing. Reporting unit: congressional district. Unrestricted.
- 41-A-18 Survey of Intentions (quarterly): Buying behavior and intentions on major household items. Reporting unit: household. Confidential; permanent; 1959 to present; 456,000 cards; 150 tapes.
- 41-A-19 Current Population Survey (monthly): Labor force status, age, sex, veteran status, education, mobility, income, and housing. Reporting unit: persons in households. Confidential; permanent; 1959 to present; 2,720,000 cards; 350 tapes.
- 41-A-20 National and State Population Estimates and Forecasts (monthly): Age, sex, births, deaths, and mobility for current or future years. Reporting unit: United States. Unrestricted; 2 years; 80,000 cards; 77 tapes.
- 41-A-21 Special Censuses (selected intervals): Age, sex, and relationship. Reporting unit: persons in households. Confidential.
- 41-A-22 Housing Inventory Change and Residential Financing (one-time survey): Components of inventory change and other housing characteristics. Reporting unit: housing units. Confidential; permanent; 1959-59; 302 tapes.
- 41-A-23 AC Summaries, Import Statistics of United States (monthly, quarterly, annual): Quantity and dollar value at reporting level. Reporting unit: summarization of import entries in terms of commodity classification (schedule A-1959, August 1963; schedule TSUSA from September 1963 to date), country of origin (schedule C) including economic class and SITC from 1963. Unrestricted; permanent; 1960-64; 175 tapes.
- 41-A-24 No. 1 Summaries, Export and Intransit Waterborne Trade of United States (annual): Quantity in pounds, value in dollars at reporting unit. Reporting unit: summarization of waterborne shippers export declaration in terms of type of vessel service. U.S. customs port of lading (schedule D), foreign port of unloading (schedule K), country of destination (schedule C), commodity (schedule S), flag vessel, engineer channel, trade area, U.S. coastal district. Confidential; permanent; 1958-63; 100 tapes.
- 41-A-25 No. 1 Summaries, Imports and Intransit Waterborne Trade of United States (annual): Quantity in pounds, value in dollars at report level. Reporting unit: invoices. Confidential; permanent; 1958-63; 150 tapes.
- 41-A-26 ACD Summary—Import Statistics of United States (monthly, quarterly, annual): Quantity of dollar value at reporting unit. Reporting unit: summarization of import entries in terms of commodity classification (schedule A 1959, August 1963, schedule TSUSA from September 1963 to date), country of origin (schedule C), customs district of entry (schedule D), traffic rate provision, type of entry (i.e., consumption or general statistics), subgroup and economic class and SITC from January 1963 to date. Unrestricted; permanent; 400 tapes.

- 41-A-27 BCD Summaries—Foreign Trade Export Air (annual): Total quantity and value, U.S.-flag carriers, quantity and value for level stated (commodity, country, district). Reporting unit: summarization of shippers export declarations of airborne shipments in terms of commodity (schedule B), country (schedule C); and district of lading (schedule D). Unrestricted; permanent; 1962-64; 30 tapes.
- 41-A-28 CQC (Cotton Quota Control) Registers—Imports (monthly): Separate registers for quantity and value, for commodity and country classification. Reporting unit: summarization of U.S. customs import entries in terms of commodity (schedule A, TSUSA and cotton quota control) and country of origin. Unrestricted; permanent; 1961 to present; 48 tapes.
- 41-A-29 ACD Summaries—Foreign Trade Import Air (annual): Total quantity and value, U.S.-flag carriers—quantity and value for level stated (commodity, country, district). Reporting unit: summarization of import entries in terms of commodity (schedule A until September 1963), country of origin (schedule C), district of unloading (schedule D). Unrestricted; permanent; 1962-64; 30 tapes.
- 41-A-30 BCD Summaries—Exports of Domestic and Foreign Merchandise (monthly): Quantity pertaining to commodity and dollar value for levels of summary stated. Reporting unit: summarization of shippers, export declarations in terms of commodity classification (schedule B), country of destination (schedule C); district of lading (schedule D). Confidential; permanent; 1963 to present; 144 tapes.
- 41-A-31 BC Summaries—Exports of Domestic and Foreign Merchandise (monthly, quarterly, annual): Quantity and value. Reporting unit: summarization of shippers export declarations in terms of commodity classification (schedule B), country of destination (schedule C). Confidential; permanent; 1958-63; 624 tapes.
- 41-A-32 Census of Manufactures (quinquennial): Plant information, number of employees, payrolls, man-hours, cost of material, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishment. Confidential; permanent; 1947; 2,500,000 cards.
- 41-A-33 Census of Manufactures and Mineral Industries (quinquennial): Plant information, number of employees, payrolls, cost of materials, inventories, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishments. Confidential; permanent; 1954; 5 million cards; 561 tapes.
- 41-A-34 Census of Manufactures and Mineral Industries (quinquennial): Plant information, number of employees, payrolls, cost of materials, inventories, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishments. Confidential; permanent; 1958; 4,500,000 cards; 752 tapes.
- 41-A-35 Economic Censuses—Retail (quinquennial): Sales, annual payroll, weekly employment, and wages. Reporting unit: retail establishments. Confidential; permanent; 1954 and 1958; 198 tapes.
- 41-A-36 Economic Censuses—Wholesale (quinquennial): Sales, annual payroll, weekly employment and wages, operating expenses, commodity lines. Reporting unit: wholesale trade establishments. Confidential; permanent; 1954 and 1958; 78 tapes.
- 41-A-37 Economic Censuses—Transportation (one-time survey): Truck information: Physical characteristics and operational aspects. Reporting unit: individuals. Confidential; 2 years; 1963; 12,000 cards; 200 tapes.
- 41-A-38 Survey of Manufactures (annual): Plant information, number of employees, payrolls, man-hours, cost of materials, inventories, capital expenditures, and products shipped. Reporting unit: manufacturing establishments. Confidential; permanent; 1949-62; 6,399,000 cards; 880 tapes.
- 41-A-39 Retail—CCBR—1 (monthly): Sales of retail firms. Reporting unit: sample of retail trade firms. Confidential; permanent; 1962 to present; 750,000 cards; 420 tapes.

- 41-A-40 Wholesale (CTR) (monthly): Sales and inventories. Reporting unit: sample of merchant wholesale establishments. Confidential; permanent; 1962 to present; 216,000 cards, 144 tapes.
- 41-A-41 County Business Pattern (annual): Number of reporting units, first quarter payroll, and number of employees. Reporting unit: individual establishments or groups of establishments. Confidential; 3 years; 1962; 250 tapes.
- 41-A-42 Economic Censuses—Services (quinquennial): Receipts, annual payroll, weekly employment, and wages. Reporting unit: service trades establishments. Confidential; permanent; 1954 and 1958; 124 tapes.
- 41-A-43 Census of Governments—Governmental Units File (quinquennial): Population or enrollment; type of government or SMSA code. Reporting unit: State, county, city, etc., school systems. Unrestricted; permanent; 1962; 3 tapes.
- 41-A-44 Census of Governments—Local Government Directory Listing (quinquennial): Population or enrollment; type of government or SMSA code. Reporting unit: counties, municipalities, townships, and school districts. Unrestricted; permanent; 1962; 140,000 cards.
- 41-A-45 Census of Governments—Employment Data (quinquennial): Employment. Reporting unit: local governments and school systems. Unrestricted; permanent; 1962; 220,000 cards; 40 tapes.
- 41-A-46 Census of Governments—Assessed Valuations (quinquennial): Description and assessed value of approximately 1 million sample pieces of property. Reporting unit: city or county taxing jurisdiction. Unrestricted; permanent; 1962; 18 tapes.
- 41-A-47 Census of Government—Finance Data (quinquennial): Revenue, expenditure, debt outstanding, cash and investment assets. Reporting unit: cities, townships, special districts, and school systems. Unrestricted; permanent; 1962; 84 tapes.
- 41-A-48 Current Survey Directory Testing (annual): Governmental unit identification and enrollment or population. Reporting unit: sample of local governmental units. Unrestricted; permanent; 1962; 1 tape.
- 41-A-49 Sample Employment Data (annual): Full-time employees and annual rate of pay. Reporting unit: local governments and school systems. Unrestricted; permanent; 1962; 30,000 cards; 4 tapes.
- 41-A-50 Retail Trade Survey (weekly): Retail sales. Reporting unit: retail establishments. Confidential; 1962 to present; 300,000 cards.
- 41-A-51 Accounts Receivable Survey (monthly): Receivables from charge accounts and installment sales. Reporting unit: sample of retail trade establishments. Confidential; permanent; 1963 to present; 240,000 cards; 384 tapes.
- 41-A-52 Retail Inventory Survey (RIS) (monthly): Inventory. Reporting unit: retail establishments and retail multiunit firms. Confidential; 39 months; 1961 to present; 72,000 cards.
- 41-A-53 Geographic Area Sales Survey (GASS) (monthly): Sales. Reporting unit: retail establishments. Confidential; permanent; 1962 to present; 128,000 cards; 448 tapes.
- 41-A-54 Service Trade Survey (monthly): Receipts for services. Reporting unit: service establishments. Confidential; 39 months; 1962 to present; 160,000 cards; 144 tapes.
- 41-A-55 Current Trade Survey (SSA Births) (quarterly): Sales and inventory. Reporting unit: wholesale establishments. Confidential; 39 months; 1961 to present; 7,500 cards.
- 41-A-56 Canned Food Survey (selected intervals): Inventory. Reporting unit: wholesale distributors and retail multiunit organizations. Confidential; 39 months; 10,800 cards.
- 41-A-57 Retail Trade Survey—Group 1 (annual): Sales and inventory. Reporting unit: retail establishments. Confidential; 39 months; 1961-63; 150,000 cards.
- 41-A-58 Retail Trade Survey—GROUP II (annual): Sales, inventory and capital expenditures. Reporting unit: retail establishments. Confidential; 39 months; 1961-63; 8,000 cards.
- 41-A-59 Capital Expenditures Survey—Wholesale Trade (quinquennial): Capital expenditures. Reporting unit: wholesale establishments. Confidential; 39 months; 1963; 28,500 cards.

- 41-A-60 Capital Expenditures Survey—Service trade (quinquennial): capital expenditure. Reporting unit: service establishments. Confidential; 39 months; 1963.
- 41-A-61 Iron and Steel Foundries, Blast Furnaces and Steel Ingot Producers (monthly): Shipments and unfilled orders for castings, production of steel castings by type of furnace and production of scrap, pig iron and iron ore by type of furnace. Reporting unit: 1,200 establishments. Confidential; 3 years; 1961-63; 116,136 cards.
- 41-A-62 Nonferrous Castings—M33E (monthly): Shipments and unfilled orders nonferrous castings by type of casting. Reporting unit: 600 establishments. Confidential; 3 years; 1961-63; 74,916 cards.
- 41-A-63 Aluminum Producers and Importers (BDSAF-122) (annual): Total receipts and shipments; shipments on ACM and rated orders; inventories of scrap, primary ingot and secondary ingot. Reporting unit: establishments. Confidential; 3 years; 1961-63; 44,496 cards.
- 41-A-64 Farm Machines and Equipment—M35A (quarterly): Production, shipments and inventory by type of farm equipment; value by product class. Reporting unit: 140 establishments. Confidential; 3 years; 1961-63; 9,324 cards.
- 41-A-65 Metalworking Machinery—M35W (quarterly): Shipments and unfilled orders (domestic and export) by type of machine, including numerical control type; units value. Reporting unit: 476 establishments. Confidential; 3 years; 1961-63; 35,208 cards.
- 41-A-66 Copper Forms and Products—Brass and Bronze Foundries (BDSAF83) (quarterly): Inventories, receipts, and shipments of copper and copper-base alloy. For brass and bronze foundries: shipments of controlled materials; also, authorized controlled materials shipments by DMS allotment number. Reporting unit: 460 establishments. Confidential; 3 years; 1961-63; 28,404 cards.
- 41-A-67 Copper Controlled Materials—Brass Mills and Copper Wire Mills (BDSAF84) (quarterly): Copper and copper-base alloy controlled materials shipments and unfilled orders; authorized controlled materials shipments by DMS allotment number. Reporting unit: 100 establishments. Confidential; 3 years; 1961-63; 7,836 cards.
- 41-A-68 Copper Forms and Products—Copper Base Powder Mills (BDSAF-574) (quarterly): Inventories, receipts, and shipments of copper and copper-base alloy; shipments of copper and copper-base alloy controlled materials and authorized controlled materials; total copper-base powder shipments by type. Confidential; 3 years; 1961-63; 1,884 cards.
- 41-A-69 Manmade Fiber, Woolen and Worsted Fabrics (monthly): Productions, stocks, and unfilled orders, gray goods and finished manmade fiber fabrics and wool apparel fabrics. Reporting unit: 200 to 250 manufacturers. Confidential; 3 years; 1961-63; 77,364 cards.
- 41-A-70 Finishing Plant Report—Broad Woven Fabrics (monthly): Gray goods inventory, finished fabrics during the period, finished goods inventory and backlog of finishing orders. Reporting unit: 130 to 150 manufacturers. Confidential; 3 years; 1961-1963; 32,436 cards.
- 41-A-71 Piece Goods Inventories and Orders (monthly): Converters, wholesalers, jobbers, and other dealers: inventories owned by the company according to location and unfilled orders for gray and finished goods. Reporting units: 200 to 225 companies. Confidential; 3 years; 1961-63; 22,896 cards.
- 41-A-72 Tufted Textile Fabrics—M22L (semiannual): Tufting machines; yarns and fabrics consumed in manufacture of auto and aircraft carpeting. Reporting unit: 1950 to 170 manufacturers. Confidential; 3 years; 1963; 702 cards.
- 41-A-73 Cotton and Linters in Public Storage and at Compresses—M22N (monthly): Raw cotton and linters inventory. Reporting unit: 1,200 to 1,220 companies. Confidential; 3 years; 1961-63; 101,592 cards.
- 41-A-74 Cotton, Manmade Fiber Staple and Linters—M22P (monthly): Consumption, stocks and spindle activity. Reporting unit: 850 to 875 companies. Confidential; 3 years; 1961-63; 135,468 cards.

- 41-A-75 Men's Apparel Cuttings and Shipments—M23B (monthly): Cuttings, shipments and value of shipments. Reporting unit: 630 to 650 manufacturers and contractors. Confidential; 3 years; 1961-63; 87,480 cards.
- 41-A-76 Women's, Misses, and Juniors' Apparel—Cuttings and Shipments—M23H (monthly): Cuttings, shipments and value of shipments. Reporting unit: 1,190 to 1,210 manufacturers and contractors. Confidential; 3 years; 1961-63; 151,272 cards.
- 41-A-77 Shoes and Slippers, Production and Shipments—M31A (monthly): Shoes and slippers except those with sole vulcanized to fabric upper. Shoes and slippers with sole vulcanized to fabric upper. Value of shipments. All other products. Reporting unit: 470 to 490 manufacturers. Confidential; 3 years; 1961-63; 114,732 cards.
- 41-A-78 Broad Fabrics (Expect knit), Woven, Nonwoven and Felts—M22T (quarterly): Loom hours, yarns consumed; looms in place and active; number of looms operating and production; and stocks of selected items. Reporting unit: 690 to 710 manufacturers. Confidential; 3 years; 1961-63; 155,184 cards.
- 41-A-79 Mattresses and Bedsprings—M25E (monthly): Quantity and value of shipments. Reporting unit: 225 manufacturing firms. Confidential; 3 years; 1961-63; 8,900 cards.
- 41-A-80 Pulp Paper and Board—M26A (monthly): Production—inventories and consumption. Reporting unit: 665 manufacturing establishments. Confidential; 3 years; 1961-63; 121,000 cards.
- 41-A-81 Inorganic Chemicals—M28A.1 (monthly): Production and inventories. Reporting unit: 495 manufacturing establishments. Confidential; 3 years; 1961-63; 51,000 cards.
- 41-A-82 Industrial Gases—M28A.2 (monthly): Production. Reporting unit: 620 manufacturing establishments. Confidential; 3 years; 1961-63; 45,000 cards.
- 41-A-83 Paint, Varnish and Lacquer—M28F (monthly): Production and sales. Reporting unit: 225 manufacturing establishments. Confidential; 3 years; 1961-63; 3,600 cards.
- 41-A-84 Rubber—M30A (monthly): Production receipts—shipments consumption—inventories. Reporting unit: 390 manufacturing establishments. Confidential; 3 years; 1961-63; 62,000 cards.
- 41-A-85 Rubber Consumption by Product Group—M30B (quarterly): Consumption. Reporting unit: 39 manufacturing establishments. Confidential; 3 years; 1961-63; 62,000 cards.
- 41-A-86 Plastic Bottles—M30E (monthly): Shipments by end use. Reporting unit: 79 manufacturing establishments. Confidential.
- 41-A-87 Oilseeds, Beans and Nuts—Report of Primary Processors—M20J (monthly): By type—quantity of seeds or beans, crushed oil production, and inventories. Reporting unit: 328 establishments. Confidential; 3 years; 1961-63; 93,024 cards.
- 41-A-88 Animal and Vegetable Fats and Oils, Inventories—M20H (monthly): End of month inventories—by types of oil. Reporting unit: 264 warehouses. Confidential; 3 years; 1961-63; 21,744 cards.
- 41-A-89 Fats and Oils—Report of Renderers—M20L (monthly): Rendered production, shipments and inventories, by type. Reporting unit: 430 establishments. Confidential; 3 years; 1961-63; 61,992 cards.
- 41-A-90 Animal and Vegetable Fats and Oils—Report of Consumers—M20M (monthly): Types and quantities of oils produced, consumed and inventories. Reporting unit: 503 establishments. Confidential; 3 years; 1961-63; 142,308 cards.
- 41-A-91 Animal and Vegetable Fats and Oils—Report of Producers and Consumers—M20N (monthly): Production, consumption and stock on hand—by type. Reporting unit: 365 establishments. Confidential; 3 years; 1961-63; 145,332 cards.
- 41-A-92 Confectionery Survey—MA20D (annually): Types of products produced, net shipments, and type of customer (to whom sold), and cost and type of materials consumed. Reporting unit: 244 manufacturers. Confidential; 3 years; 1961-63; 16,047 cards.
- 41-A-93 Formula Feeds for Poultry and Livestock—MA20E (annual): Production and shipments by type. Reporting unit: 3,077 manufacturers. Confidential; 3 years; 1961-63; 16,698 cards.

- 41-A-94 Salad Dressing, Mayonnaise and Related Products—MA20F (annual): Production by type and size of container, value of shipments, number of brands and oils consumed. Reporting unit: 135 establishments. Confidential; 3 years; 1961-63; 17,475 cards.
- 41-A-95 Fats and Oils—Report of Renderers—MA20L (annual): Production, consumption, and inventories. Reporting unit: 1,170 establishments. Confidential; 3 years; 1961-63; 10,182 cards.
- 41-A-96 Animal and Vegetable Fats and Oils—Report of Producers and Consumers—MA20N (annual): Production, consumption and inventories by type. Reporting unit: 1,142 establishments. Confidential; 3 years; 1961-63; 16,062 cards.
- 41-A-97 Woolen and Worsted Machinery Activity—MA22E (annual): Spindles in place and active the last full working day of year, and worsted combs and top to tow converters in place and active on last workday of year. Reporting unit: 350 establishments. Confidential; 3 years; 1961-63; 1,518 cards.
- 41-A-98 Yarn Production—MA22F (annual): Production data for yarn spun from staple, tow and uncut top and for stretch, textured, crimped or bulked filament yarns. Reporting unit: 1,000 to 1,200 companies. Confidential; 3 years; 1961-63; 12,000 cards.
- 41-A-99 Narrow Fabrics—MA22G (annual): Production; materials consumed; narrow fabric machinery in place at end of year; value of shipments. Reporting unit: 300 establishments. Confidential; 3 years; 1962-63; 5,686 cards.
- 41-A-100 Knit Cloth for Sale—MA22K (annual): Shipments of knit cloth for sale; yarns consumed in production. Reporting unit: 300 to 400 establishments producing knit cloth for sale. Confidential; 3 years; 1961-63; 5,949 cards.
- 41-A-101 Tufted Textile Products—MA22L (annual): Tufting machinery in place at end of year; automobile and aircraft carpeting, quality and value of shipments and yarns consumed. Reporting unit: 200 manufacturers. Confidential; 3 years; 1961-63; 1,053 cards.
- 41-A-102 Stocks of Wool and Related Fibers—MA22M (annual): Stock of foreign and domestic wool and stocks of related fibers and tops and noil of January 1 of each year. Reporting unit: 500 to 600 establishments. Confidential; 3 years; 1961-63; 6,000 cards.
- 41-A-103 Cotton and Linters in Public Storage and at Compresses—MA22N (annual): Stocks on hand on July 31 of each year; material in transit on July 31 and destroyed during the year. Reporting unit: 150 to 250 establishments that do not report on the monthly M22N. Confidential; 3 years; 1961-63; 450 cards.
- 41-A-104 Cotton, Manmade Fiber Staple, and Linters—MA22P (annual): Consumption during year and stocks at end of year; spindles in place and active and spindle hours operated during the year; destroyed during season. Reporting unit: 250 to 300 establishments that do not report on the monthly M22P. Confidential; 3 years; 1961-63; 1,878 cards.
- 41-A-105 Rugs, Carpets and carpeting—MA22Q (annual): Shipments and interplant transfers; yarns and fabrics consumed in the manufacturing of specified types of carpets and rugs; machinery in place at yearend. Reporting unit: 100 to 125 establishments. Confidential; 3 years; 1962-63; 700 cards.
- 41-A-106 Cotton, Silk and Manmade Fiber Woven Goods Finished—MA22S (annual): Fabrics finished by end use, for crease resistance or wash-wear properties and against military contracts. Reporting unit: 350 to 400 finishing plants. Confidential; 3 years; 1961-63; 14,436 cards.
- 41-A-107 Broad Woven Goods, Cotton, Wool, Silk and Manmade Fiber—MA22T (annual): Production of wool woven fabrics, number of looms operating on last working day; looms in place and active by shifts; loom hours; yarns consumed; stocks. Reporting unit: 200 to 225 established not reporting on the quarterly MA22T. Confidential; 3 years; 1961-63; 3,414 cards.
- 41-A-108 Apparel Survey—MA23A (annual): Production and value of shipments, knit yarns consumed. Reporting unit: 6,000 to 7,000 manufacturers and jobbers of apparel. Confidential; 3 years; 1961-63; 167,385 cards.



- 41-A-109 Brassiers, Corsets and Allied Garments Shipments—MA23J (annual): Quantity and value of shipments. Reporting unit: 200 to 250 producers. Confidential; 3 years; 1961-63; 6,809 cards.
- 41-A-110 Brassiers, Corsets and Allied Garments—Distribution of Sales—MA23J (annual): Distribution of manufacturers sales by class of customer. Reporting unit: 175 to 200 companies reporting on the MA23J which had a value of shipments of over \$250,000 the previous year. Confidential; 3 years; 1961-63; 2,826 cards.
- 41-A-111 Hardwood Plywood—MA24F (annual): Products and shipments. Reporting unit: 265 manufacturing establishments. Confidential; 3 years; 1961-63; 3,600 cards.
- 41-A-112 Softwood Plywood—MA24H (annual): Production and consumption. Reporting unit: 153 manufacturing establishments. Confidential; 3 years; 1961-63; 2,850 cards.
- 41-A-113 Softwood Veneer—MA24K (annual): Production and consumption. Reporting unit: 88 manufacturing establishments. Confidential; 3 years; 1961-63; 2,450 cards.
- 41-A-114 Mattresses and Bedsprings—MA25E (annual): Quantity and value of shipments. Reporting unit: 30 to 40 manufacturing establishments. Confidential; 3 years; 1961-63.
- 41-A-115 Pulp, Paper and Board—MA26C (annual): Receipts, production, inventories, consumption, shipments. Reporting unit: 717 manufacturing establishments. Confidential; 3 years; 1961-63; 37,500 cards.
- 41-A-116 Converted Flexible Packaging Products—MA26F (annual): Quantity and value of shipments. Reporting unit: 125 manufacturing establishments. Confidential.
- 41-A-117 Sulfuric Acid—MA28B (annual): Production 1-1 quantity and value of shipments. Reporting unit: 219 manufacturing establishments. Confidential; 3 years; 1961-63; 900 cards.
- 41-A-118 Shipments and Production of Inorganic Chemicals—MA28E.1 (annual): Production, consumption, quantity, and value of shipments. Reporting unit: 995 manufacturing establishments. Confidential; 3 years; 1961-63; 3,300 cards.
- 41-A-119 Shipments and Production of Industrial Gases—MA28E.2 (annual): Production, shipments, and consumption. Reporting unit: 696 manufacturing establishments. Confidential; 3 years; 1961-63; 5,300 cards.
- 41-A-120 Pharmaceutical Preparations, Except Biologicals—MA28G (annual): Value of shipments. Reporting unit: 1,100 manufacturing establishments. Confidential; 3 years; 1961-63; 45,000 cards.
- 41-A-121 Plastic Products—MA30D (annual): Quantity and value of shipments and consumption. Reporting unit: 2,846 manufacturing establishments. Confidential; 3 years; 1961-63; 65,000 cards.
- 41-A-122 Plastic Bottles—MA30E (annual): Shipments of plastic bottles by end use. Reporting unit: 60 manufacturing establishments. Confidential.
- 41-A-123 Shoe and Slipper Production and Shipments—MA31A (annual): Production and shipments of shoes and slippers and value of shipments. Reporting unit: 450 to 500 establishments that do not report on the monthly M31A. Confidential; 3 years; 1961-63; 18,960.
- 41-A-124 Steel Mill Products—MA33B (annual): Quantity of receipts, production and products consumed in manufacture; quantity and value of interplant transfers, carbon steel, alloy, steel, stainless steel and conversion steel. Reporting unit: 483 establishments. Confidential; 3 years; 1961-63; 11,946 cards.
- 41-A-125 Steel Power Boilers—MA34G (annual): Orders booked by type of boiler. Reporting unit: 100 establishments, manufacturing. Confidential; 3 years; 1961-63.
- 41-A-126 Heating and Cooking Equipment (Except Electric)—MA34N (annual): Quantity and value of shipments and yearend inventory for gas heating stoves, gas ranges, oil burners, furnaces, stokers, and water heaters. Reporting unit: 450 establishments. Confidential; 3 years; 1961-63; 7,263 cards.

- 41-A-127 Farm Machines and Equipment—MA35A (annual): Farming machinery—production, number and value of shipments (domestic and export). Reporting unit: 1,100 establishments. Confidential; 3 years; 1961-63; 16,119 cards.
- 41-A-128 Construction Machinery—MA35D (annual): Shipments (domestic and export) in number units and value. Reporting unit: 88 establishments. Confidential; 3 years; 1961-63; 2,400 cards.
- 41-A-129 Mining Machinery—MA35F (annual): Total shipments (domestic and export)—quantity and value. Reporting unit: 190 establishments. Confidential; 3 years; 1961-63; 1,746 cards.
- 41-A-130 Internal Combustion Engines—MA35L (annual): Quantity and value by model number and engine specifications of shipments to other companies or transferred to other plants of the same company; engines converted from engines received from other plants. Reporting unit: 98 establishments. Confidential; 3 years; 1961-63; 5,868 cards.
- 41-A-131 Air-Conditioning and Refrigeration Equipment—MA35M (annual): Total shipments in number of units and value in dollars. Reporting unit: 254 establishments. Confidential; 3 years; 1961-63; 10,014 cards.
- 41-A-132 Pumps and Compressors—MA35P (annual): Pumps, compressors (shipments in number of units, and value of driven units). Reporting unit: 442 establishments. Confidential; 3 years; 1961-63; 8,298 cards.
- 41-A-133 Office, Computing and Accounting Machines—MA35R (annual): Cash registers and data processing machines; typewriters; duplicating machines; number shipped; f.o.b. plant value, and retail list price. Reporting unit: 220 establishments. Confidential; 3 years; 1961-63; 4,377 cards.
- 41-A-134 Metal Working Machinery: Metal Cutting and Metal Forming Types—MA35W (annual): Shipments and unfilled orders (domestic and export) by type of machine, including numerical control type. Reporting unit: 476 establishments. Confidential; 3 years; 1961-63; 35,208 cards.
- 41-A-135 Switchgear, Switchboard Apparatus, Relays, and Industrial Controls—MA36A (annual): Value of shipments; switchgear, circuit breakers, low-voltage panelboards and distribution boards; fuses and fuse equipment under 2,300 volts; circuit relays. Reporting unit: 573 establishments. Confidential; 3 years; 1961-63; 7,452 cards.
- 41-A-136 Electric Housewares and Fans—MA36E (annual): Total shipments, quantity and value, of electric fans (non-industrial); electric razors, and small household electric cooking and heating appliances. Reporting unit: 319 establishments. Confidential; 3 years; 1961-63; 2,616 cards.
- 41-A-137 Motors and Generators—MA36H (annual): Fractional horsepower motors; integral horsepower motors and generators other than for land transportation equipment; land transportation motors, generators, and control equipment and parts; prime mover generator sets; rotating equipment. Number and value of shipments and interplant transfers. Reporting unit: 279 establishments. Confidential; 3 years; 1961-63; 5,223 cards.
- 41-A-138 Wiring Devices and Supplies—MA36K (annual): Current carrying and noncurrent carrying devices—quantity and value. Reporting unit: 383 establishments. Confidential; 3 years; 1961-63; 3,564 cards.
- 41-A-139 Lighting Fixtures—MA36L (annual): Value of shipments—electric lighting fixtures (residential, commercial, and institutional types); vehicular lighting equipment; outdoor lighting equipment including components and parts. Reporting unit: 771 establishments. Confidential; 3 years; 1961-63; 8,295 cards.
- 41-A-140 Selected Electronic and Associated Products—MA36N (annual): Shipments of electronic component parts, equipment and systems. Reporting unit: 197 establishments. Confidential; 3 years; 1961-63; 22,464 cards.

- 41-A-141 Selected Instruments and Related Products—MA3SB (annual): Value of shipments—test, measuring, and analyzing equipment for electronic and electrical circuits; industrial process instruments; selected analytical instruments and equipment. Reporting unit: 1,541 establishments. Confidential; 3 years; 1961-63; 5,472 cards.
- 41-A-142 Atomic Energy Products and Services—MA38Q (annual): Nuclear reactors; reactor components and equipment—value of shipments. Reporting unit: 300 establishments. Confidential; 3 years; 1961-63; 1,497 cards.
- 41-A-143 Particle Board—BDSAF582 (annual): Production. Reporting unit: 58 manufacturing establishments. Confidential; 3 years; 1961-63; 1,000 cards.
- 41-A-144 Lumber Survey—MA24T (annual): Production and stocks by species. Reporting unit: 4,500 sawmills. Confidential; permanent; 1954 to present; 2,700 cards; 500 tapes.
- 41-A-145 Manufacturers' Shipments, Inventories and Orders Survey—M3 (monthly): Value of shipments, new orders, unfilled orders, total inventory—materials, and supplies goods in process, and finished goods. Reporting unit: single unit; single unit and divisional manufacturing establishments. Confidential; permanent; 1960-64; 252,000 cards; 48 tapes.
- 41-A-146 Survey of Research and Development—RD1 and RD2 (annual): Company funds expended resulting from research and development, company net sales, etc. Reporting unit: 3,000 companies collecting and compiling R. & D. data for the National Science Foundation. Confidential; permanent; 1957-63; 21,000 cards; 10 tapes.
- 41-A-147 Nonpermit Construction Starts (monthly): Number of housing units started in selected nonpermit primary sampling units (PSU's) inside or outside a segment date of start and intent of building (i.e., for sale, rent, etc.). Reporting unit: 800-200 owners or contractors. Confidential; 12 years; 1959 to present; 50,000 cards.
- 41-A-148 Reports of Building Permits Issued (monthly): Number of residential buildings, housing units, and valuation by type of structure and total number of buildings, and valuation for each non-residential item reported. Reporting unit: 4,300 building or zoning officials of selected (sample) permit issuing municipalities, towns, cities, or villages. Unrestricted; 12 years; 1955 to present; 1,040,000 cards.
- 41-A-149 Residential Building Permit Use Survey (monthly): Number of units started in selected permit issuing places, date of start, type of structure (i.e., 1-family, etc.) and intent of building (i.e., for sale, rent or exclusive use). Reporting unit: 5,000-6,000 permit offices. Confidential; 12 years; 1960 to present; 300,000 cards.
- 41-A-150 Construction Progress Report Survey (monthly): Type of construction, total cost (i.e., earnings, materials, etc.) and date of start. Reporting unit: 30,000 new construction owners or contractors via (F.2. Dodge data collecting agency by questionnaires. Confidential; 12 years; 1960 to present; 240,000 cards; 56 tapes.
- 41-A-151 Housing Sales Survey (monthly): Number of units sold and unsold, stage of construction (i.e., started, completed, or not started) at time of sale, sale price, and type of finance (i.e., FHA, VA, conventional). Reporting unit: 6,000-6,500 permit offices and camera/contractors of building sites in nonpermit places via a field enumerator. Confidential; 12 years; 1961 to present; 288,000 cards.
- 41-A-152 Report of Building permits Issued (annual): Number of buildings, housing units, and valuation of type of structure, private and public. Reporting unit: 8,000 building or zoning officials of permit-issuing places other than those canvassed monthly. Unrestricted; 12 years; 1960-63; 175,000 cards.

## BUREAU OF INTERNATIONAL COMMERCE

- 41-B-1 World Trade Director Reports File, Japan (selected intervals): Describes foreign company, products handled, manufacturer, size, reputation, capital, annual turnover and other pertinent facts. Reporting unit: foreign posts, based on their contacts with foreign companies. Unrestricted; 5 years; 22,000 cards.

- 41-B-2 American Traders Identification File (annual): Investment interest, number of employees, annual sales import banks ref., SIC nos., address, principal officer. Reporting unit: American companies on form 1A-57. Confidential; 5 years; 12,000 cards.
- 41-B-3 Dun & Bradstreet "Million Dollar Directory" Card Check (one-time survey): Number of employees, annual sales, capital, principal officer and SIC number. Reporting unit: Dun & Bradstreet based on their contacts with American companies. Confidential; 3 years; 60,000 cards.
- 41-B-4 Tariff and Trade Negotiation (Gatt) Committees (one-time survey): Tariff information and import statistics. Reporting unit: data based on government publications. Confidential; permanent; 1961-63; 250,000 cards.
- 41-B-5 Corporation Income Tax Returns (annual): Balance sheet and income statement items and asset size classes. Reporting unit: U.S. companies via Internal Revenue Service. Confidential; Permanent; 1959-60; 112 tapes.
- 41-B-6 Trade of OECD Countries and Japan (annual): Imports and exports between 20 foreign countries. Reporting unit: OECD and United Nations—data based on country reports. Unrestricted; permanent; 1961-62; 20 tapes.
- 41-B-7 U.S. Exports of Domestic and Foreign Merchandise (annual): Quantity and value by schedule B numbers for countries of destination. Reporting unit: exporters' reports to collectors of custom on shipments abroad. Unrestricted; permanent; 1962; 10 tapes.
- 41-B-8 World Trade Directory Reports File, Japan (revised) (annual): Describes foreign company, products handled, manufacturer, size, reputation, capital, annual turnover and other pertinent facts. Reporting unit: foreign posts, based on their contacts with foreign companies. Unrestricted; 5 years; 75,000 cards.
- 41-B-9 U.S. Imports of Merchandise (annual): Quantity and value by schedule A by countries of origin. Reporting unit: importers report to collectors of customs on shipments arriving in the United States. Unrestricted; permanent; 1962; three tapes.

## BUREAU OF MINES

- 42-1 Mine and Quarry Survey (annual): Footage of exploration and development, quality of ore mines and mining methods. Reporting unit: quantities shortages, long tons. Confidential; permanent; 1958-63; 275,000 cards; 75 tapes.
- 42-2 Clay (annual): Reports on quantity and value used and quantity and value sold, total quantity used and sold and value total used and sold. Reporting unit: producers of clay. Confidential; 1 year; 1962-63; 10,000 cards.
- 42-3 Sand and Gravel Canvass (annual): Production figures of various classes S&G by government and commercial producers and users. Reporting unit: sand and gravel producers & users (commercial and non-commercial). Confidential; permanent; 1961-63; 60,000 cards, 7 tapes.
- 42-4 Water Canvass (quinquennial): Water consumption, production, treatment, source of water recirculation. Reporting unit: various mineral producers. Unrestricted; permanent; 1962; 86,000 cards; 16 tapes.
- 42-5 Iron and Steel (scrap iron) (monthly): Shipments order for casting, production of steel castings and ingots by type of furnace, consumption of scrap, pig iron, iron ore, stocks, production, receipts and consumption by grades of scrap. Reporting unit: users of scrap iron. Unrestricted; 1 year; 1963-64; 84,000 cards.
- 42-6 Bituminous and anthracite coal (monthly): Employment, number of operators, days active, days lost, number of injuries; man-hours, man-days and production. Unrestricted; 1 year; 25,000 cards.
- 42-7 Bituminous and Anthracite Coal Canvass (annual): Annual bituminous and anthracite survey covering employment, days and hours worked, disabling injuries and production. Reporting unit: coal producers. Unrestricted; 10 years; 245,000 cards.

- 42-8 Stone Quarries Canvass (annual): Survey of the industry covering number of operations, employment, man-hours worked, days active, length of shift, disabling injuries, including fatalities. Reporting unit: quarry operators. Confidential; 8 years; 80,000 cards.
- 42-9 Petroleum Injury Experience (annual): Data from oil companies, including oil, and gas-well drilling contractors, oil- and gas-field service contractors, and natural gas companies. Reporting unit: petroleum producers. Unrestricted; 8 years; 60,000 cards.
- 42-10 Sand and Gravel Injury Experience (annual): Employment, man shifts, man-hours, injuries, days lost, degree of injuries and number of plants. Reporting unit: sand and gravel operators and users. Unrestricted; 6 years; 66,000 cards.
- 42-11 Metal Industries Including Placer (annual): Same as nonmetal industry. Reporting unit: metal operators and producers. Confidential; 8 years; 70,000 cards.
- 42-12 Nonmetal Industries (except quarries and coal) (annual): Number of men employed, days and hours worked, number of mines, injuries, days lost, degree of injury, nature and part of body. Reporting unit: nonmetal operators and producers. Confidential; 8 years; 71,000 cards.
- 42-13 Manufacturers and Coal Retailers (monthly): Tons on hand, tons received during month, tons consumed and tons on hand at end of month. Reporting unit: consumers of coal. Unrestricted; 2 years; 24,000 cards.

## DEPARTMENT OF LABOR

## BUREAU OF EMPLOYMENT SECURITY

- 44-A-1 Teuc (Temporary Extended Unemployment Compensation) family characteristics study (one-time survey): Personal characteristics (age, sex, number in household, etc.), economic characteristics (industry, occupation, work history), and unemployment insurance experience (weekly benefit, duration of benefits, etc.). Reporting unit: individuals-claimants for Teuc. Confidential; 3 years; 1961-62; 230,000 cards.
- 44-A-2 ES-212 nonagricultural placements, standard metropolitan areas (monthly): Major occupation group—total, veteran, under 22, 45 and over. Industry division and 2-digit manufacturing—total, veteran, 45 and over. Reporting unit: 65 selected SMSA reports. Unrestricted; permanent; 1962-63; 73,000 cards.
- 44-A-3 ES-212 nonagricultural placements, State summaries (monthly): Major occupation group—total, female, veteran, age under 22, age 45 and over. Industry division and 2-digit manufacturing—total, female, veteran, age 45 and over. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1957-63; 90,900 cards.
- 44-A-4 ES-211 employer information transcript (monthly): Employment past, current, future; turnover and placements; current hours and wage rate. Reporting unit: individual firms for selected industries. Confidential; permanent; 1959-63; 32,000 cards.
- 44-A-5 ES-209 supplement—service to selected age groups (monthly): New applications, active file, initial counsel interview, nonagricultural placements—total, female, age under 22, age 45-64. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1962-63; 5,400 cards.
- 44-A-6 ES-209 local office activities, standard metropolitan areas (monthly): Applicant services, employer services, nonagricultural placement activities, claims-taking activities—total, veteran. Reporting unit: 65 selected SMSA reports. Unrestricted; permanent; 1962-63; 54,100 cards.
- 44-A-7 ES-209 local office activities, State summaries (monthly): Applicant services, employer services, nonagricultural placement activities, agricultural placement activities, claims activities—total, female and veteran. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1962-63; 54,100 cards.

- 44-A-8 ES-202 employment and wages of workers covered by UI and UCFE laws (quarterly): Reporting units, 3-month employment, total wages, taxable wages and contributions. Reporting unit: State summary reports (51 State agencies). Confidential; permanent; 1958-63; 1,810,000 cards; 24 tapes.
- 44-A-9 ES-203 characteristics of the insured unemployed (monthly): Age and sex of the insured employed; sex and duration by industry; industry and occupation. Reporting unit: insured unemployed individuals. Unrestricted; permanent; 1958-63; 540,000 cards.

## BUREAU OF LABOR STATISTICS

- 44-B-1 Wholesale price index (monthly): Wholesale price and discounts for specified products. Reporting unit: producers of commodities for primary markets. Confidential; 5 years; 1959-present; 310,000 cards.
- 44-B-2 Comprehensive housing unit survey (decennial): Type of housing unit, occupancy, tenure, rent or value, year built, rooms, persons, equipment, income level. Reporting unit: individual housing units; selected cities. Confidential; 5 years; 1959-60-63; 500,000 cards.
- 44-B-3 Consumer price index—foods (monthly): Retail price of 120 food and household supply items. Reporting unit: grocery stores in selected urban areas. Confidential; 4 years; 1960-present; 5,100,000 cards.
- 44-B-4 Consumer price index—home ownership (monthly): Sales price, size (square feet), whether previously occupied, year built, site value. Reporting unit: data supplied by FHA; derived from insured mortgage file. Confidential; permanent; 1958-present; 1,850,000 cards; 55 tapes.
- 44-B-5 Consumer price index—rents (biennial): Monthly rental, date of rent change, change in characteristics of housing unit and rental terms. Reporting unit: individual housing units in selected urban areas. Confidential; 2 years; 1962-present; 200,000 cards.
- 44-B-6 Consumer expenditure survey (decennial): Detailed accounting of household receipts and disbursements, and family characteristics. Reporting unit: individual households. Confidential; 10 years; 1959-60-63; 5,500,000 cards; 1,200 tapes.
- 44-B-7 National survey of professional, administrative and technical pay—(annual): Frequency and earnings of selected professional, administrative and technical personnel. Reporting unit: 1,700 establishments in manufacturing, public utilities, wholesale trade, retail trade, and engineering services. Confidential; permanent; 450,000 cards, 3 tapes.
- 44-B-8 Earnings distribution survey—minimum wage impact (selected intervals): Hours and straight-time earnings for individual employees, nonoccupational, e.g., all production or nonsupervisory. Reporting unit: establishments in broad industry groups—retail trade, manufacturing, wholesale trade. Confidential; 5 years; 1958-62; 790,000 cards.
- 44-B-9 Union wage scales in the building trades, printing trades, local transit and local trucking in 64 cities (annual): Minimum union wage scale and maximum hours for selected occupations. Reporting unit: union locals. Unrestricted; 3 years; 30,000 cards.
- 44-B-10 Supplementary employee remuneration surveys—manufacturing and broad industry groups (triennial): Establishment expenditure on selected supplementary fringe items. Reporting unit: establishments. Confidential; 6 years; 1959-62; 200,000 cards; 2 tapes.
- 44-B-11 Strikes and lockouts (monthly): Number of stoppages, workers and man days idle involving six or more workers in excess of one shift. Reporting unit: newspaper clipping service. Unrestricted; permanent; 200,000 cards.
- 44-B-12 Industry wage occupational studies (annual, triennial, quinquennial): Straight-time earnings for selected representative occupations and supplementary benefits. Reporting unit: establishments, primarily manufacturing, on nationwide and area basis. Confidential; 5 years; 40,000 cards.

- 44-B-13 Wages and related benefits—82 labor market areas (annual): Hours and straight-time earnings for selected cross-industry jobs and supplementary establishment benefits. Reporting unit: establishments in manufacturing wholesale trade, retail trade, finance and insurance, and services. Confidential; 5 years; 2 million cards; 10 tapes.
- 44-B-14 Older worker performance study (one-time survey): Production reports for selected weeks for selected individuals (mail sorters). Reporting unit: U.S. postal service. Unrestricted; 4 years; 1962; 24,000 cards.
- 44-B-15 Seasonally adjusted employment and average weekly hours (annual): Employment and average weekly hours—50 series. Reporting unit: establishments and households. Unrestricted; permanent; 1,000 cards.
- 44-B-16 Average annual rates of change (regression coefficient) (annual): Output per man-hour, output, man-hours, employment, and unit labor costs. Reporting unit: data from labor force and household reports. Unrestricted; permanent; 2,000 cards.
- 44-B-17 Master address tape, maritime survey (quarterly): Name and address of employers subject to Longshoremen's and Harbor Workers' Compensation Act. Confidential; 1 tape.
- 44-B-18 Injury rates in manufacturing (quarterly): Employees, hours, number of work injuries. Reporting unit: establishments. Confidential; 2 years; 1961-64; 650,000 cards.
- 44-B-19 Injuries and causes of accidents to seamen (selected intervals): Description of injured, location of accident, description of injuries, causes. Reporting unit: individual accident. Confidential; 2 years; 1954-61; 7,500 cards.
- 44-B-20 Injury rates by industry (annual): Employees, hours, injury distribution and time charges. Reporting unit: establish. Confidential; 5 years; 1958-62; 500,000 cards.
- 44-B-21 Injuries and accidents to longshoremen and harbor workers (quarterly, annually): Deaths, other injuries, and hours worked, by port. Reporting unit: employers subject to Longshoremen's and Harbor Workers' Compensation Act. Confidential; 2 years; 1962-present; 50,000 cards; 1 tape.
- 44-B-22 Master registry of work-injury reporters (annual): Name; State, industry, reporter codes; city location. Reporting unit: establishment. Confidential; permanent; 1962; 1 tape.
- 44-B-23 Work injuries and work-injury rates in hospitals (one-time survey): By department: average number of employees, total man-hours. For individual injured employees; kind of injury, nature of injury, part of body injured, occupation. Reporting unit: individual hospitals. Confidential; 2 years; 1963; 2 tapes.
- 44-B-24 Injuries and accident causes (one-time survey): Occupation, extent of disability, severity, nature of injury, accident type, hazardous condition, agency of accident, unsafe act, activity of injured at time of accident. Reporting unit: individual injuries in establishments in selected industries. Confidential; 2 years; 1955-60; 10,200 cards.
- 44-B-25 Work injuries experienced by minors (one-time survey): Industry in which employed, activity of injured, nature of injury, accident type. Reporting unit: individual reports transcribed from workman's compensation files of eight States. Confidential; 2 years; 1959; 32,000 cards.
- 44-B-26 Work injuries and work-injury rates (one-time survey): By department: average number of employees, total man-hours. For individual injury employees: kinds of injuries, nature of injury, occupation, part of body injured. Reporting unit: establishments in selected industries. Confidential; 2 years; 1960-61; 104,000 cards.
- 44-B-27 Survey of industry employment, payroll and hours (monthly): All employees, women, production workers, and average hourly earnings, average weekly earnings, hours, and overtime for production workers. Reporting unit: Individual establishment. Confidential; permanent; 1957-present; 100 tapes.

- 44-B-28 Survey of industry labor turnover (monthly): Total separation, quits, layoffs, total accessions, new hires. Reporting unit: individual establishment. Confidential; permanent; 1956-present; 2,500,000 cards; 3 tapes.
- 44-B-29 National survey of scientific and technical personnel in industry (annual): Employment of approximately 20 scientific and technical occupations by function (research and development, basic research, administration, etc.). Reporting unit: company and establishment. Confidential; permanent; 1959-64; 200,000 cards; 8 tapes.
- 44-B-30 Estimates of labor force characteristics from current population survey (monthly): Employment status of noninstitutional population 14 years and older by demographic characteristics. Reporting unit: household. Confidential; permanent; 1947-present; 7 tapes.

## OTHER BUREAUS AND DIVISIONS

- 44-D-1 Applications for learner certificates (quarterly): Learners authorized. Reporting unit: applications. Confidential; 1 year; 2,000 cards.
- 44-D-2 Minors illegally employed in agriculture (semiannual): Minors illegally employed, age and school grade. Reporting unit: investigated farms. Confidential; 2 years; 14,000 cards.
- 44-D-3 Minors illegally employed in industries other than agriculture (semiannual): Minors illegally employed, age and hazardous order violated. Reporting unit: investigated establishments. Confidential; 2 years; 11,000 cards.
- 44-D-4 Learner investigations (quarterly): Learner in violation of certificate. Reporting unit: investigated establishments. Confidential; 2 years; 400 cards.
- 44-D-5 Reports on investigation findings (monthly): Employees underpaid, underpayments disclosed, minors illegally employed. Reporting unit: WHPC investigated establishments. Confidential; 2 years; 110,000 cards.
- 44-D-6 Register of all welfare and pension plans filed under the WPPDA (monthly): Name and address of plan; basic characteristics; types of employees; industry; benefits provided; type of administration; State location; type of funding; etc. Reporting unit: individual welfare and pension benefit plans. Unrestricted; permanent; 1959-present; 14 tapes.
- 44-D-7 Financial data of welfare and pension plans (one-time survey): Financial data; contributions, benefits paid; assets, by type; amount of insurance premiums; distributions by size, class; etc. Reporting unit: individual welfare and pension plans filing financial reports. Unrestricted; permanent; 1959; 110,000 cards.
- 44-D-8 Financial data of welfare and pension plans (annual): Financial data; contributions; benefits paid; assets by type; distributions by value of assets; amount of insurance premiums; distributions by size, class; etc. Reporting unit: individual welfare and pension plans filing financial reports. Unrestricted; permanent; 1960-63; 16 tapes.
- 44-D-9 Administrative costs of welfare and pension plans (one-time survey): Administrative cost data; total receipts; salary costs; fees, rents, interest; other administrative costs; etc. Reporting unit: sample (approximately 4,000) of individual welfare and pension plans reporting under the WPPDA. Unrestricted; permanent; 1962; 2 tapes.
- 44-D-10 Names and address of reporting unions (monthly): Name of president or financial officer of union and mailing address of union. Reporting unit: individual labor union reporting under the provisions of the Labor-Management Reporting and Disclosure Act. Unrestricted; permanent; 1959-present; 52,000 cards; 4 tapes.
- 44-D-11 Register of reporting labor unions (monthly): City and State location, union affiliation, unit designation. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1959-64; 6 tapes.

- 44-D-12 Characteristics of labor unions (monthly): Locations; union affiliation; unit designation; dues and fees; election rules; etc. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1959-present; 52,000 cards; 2 tapes.
- 44-D-13 Financial data for unions (selected intervals): Receipts, disbursements, assets by type. Reporting unit: individual unions reporting under LMRDA. Unrestricted; permanent; 1959-60; 208,000 cards.
- 44-D-14 Financial data for unions (annual): Receipts and disbursements by type, assets by type, etc. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1962-63; 4 tapes.

## DEPARTMENT OF THE TREASURY

## INTERNAL REVENUE SERVICE

- 48-1 Statistics of income—Individual income tax returns (annual): Sources of income, adjusted gross income, itemized deductions, exemptions, taxable income, income tax, business income, and deduction items. Reporting unit: stratified sample of income tax returns of individuals (approximately 500,000). Confidential; 3 years; 1962-62; 2,503 tapes.
- 48-2 Statistics of income—Sales of capital assets, individuals (one-time survey): Gross sales price, depreciation, cost, gain or loss, period held, and type of capital asset for each transaction; adjusted gross income and selected income items from return. Reporting unit: all returns of individuals reporting sales of capital assets included in the statistics of income sample for individual income tax returns (approximately 155,000). Confidential; 3 years; 1962; 34 tapes.
- 48-3 Statistics of income—Corporation income tax returns (annual): Assets, liabilities, receipts, deductions, profits, income tax, and tax-related items, distributions to stockholders, and industry. Reporting unit: stratified sample of income tax returns of corporations (approximately 170,000). Confidential; 3 years; 1960-62; 1,475 cards.
- 48-4 Statistics of income—Foreign tax credit study (annual): Foreign taxable income, foreign taxes, dividends received, foreign tax credit and related items, industry, country, and total assets. Reporting unit: all returns 1118 attached included in the statistics of income sample for corporation income tax returns (approximately 4,000). Confidential; 3 years; 1961-62; 21 tapes.
- 48-5 Statistics of income—Controlled foreign corporation study (one-time survey): Foreign receipts, profits, taxes, total assets, dividends paid, country, and industry. Reporting unit: all returns with form 2952 attached included in the statistics of income sample for corporation income tax returns (approximately 11,000). Confidential; 3 years; 1962; six tapes.
- 48-6 Statistics of income—U.S. business tax returns, sole proprietorships (annual): Income and deduction items and industry. Reporting unit: stratified sample of individual income tax returns—sole proprietorships (approximately 220,000). Confidential; 3 years; 1960-62; 754 tapes.
- 48-7 Statistics of income—U.S. business tax returns, partnership returns (annual): Income and deduction items, balance sheet data for alternate years, and industry. Reporting unit: stratified sample of partnership returns of income (approximately 75,000). Confidential; 3 years; 1960-62; 232 tapes.
- 48-8 Statistics of income—Estate tax returns (biennial): Gross estate, deductions, exemptions, estate tax, tax credits, types of property, executor's commission, attorney's fees, and funeral expenses. Reporting unit: all estate tax returns (approximately 80,000). Confidential; 3 years; 1962; four tapes.
- 48-9 Statistics of income—Fiduciary income tax returns (biennial): Sources of income, deductions, exemptions, and tax items. Reporting unit: stratified sample of fiduciary income tax returns (approximately 51,000). Confidential; 3 years; 1962; two tapes.

- 48-10 Statistics of income—Gift tax returns (biennial): Total gifts, exclusions, deductions, specific exemption, and gift tax. Reporting unit: stratified sample of gift tax returns (approximately 34,000). Confidential; 3 years; 1962; two tapes.
- 48-11 Tax model for individual income tax returns (annual): Income, deductions, exemptions, and tax information. Reporting unit: subsample of statistics of income sample for individual tax returns. Confidential; permanent; 1960-62; 316 tapes.
- 48-12 Tax model for corporation income tax returns (annual): Income deductions, exemptions and tax information. Reporting unit: subsample of statistics of income sample for corporation income tax returns. Confidential; permanent; 1962; 10 tapes.
- 48-13 Tax model for partnership returns (annual): Income and deduction items. Reporting unit: subsample of statistics of income sample for partnerships. Confidential; permanent; 1962; five tapes.
- 48-14 Reporting characteristics of taxpayers (one-time survey): Selected data on excise taxes and extent to which taxpayer fills out return. Reporting unit: stratified sample of form 720 (approximately 58,000). Confidential; not kept; 1963.

## DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

## OFFICE OF EDUCATION

- 51-1 School Centered Physical Fitness Programs (annual): Pupils engaged in planned program of physical activity, those passing tests of physical fitness and those who had physical exams. Reporting unit: elementary and secondary schools. Unrestricted; 2 years; 1962-63; 4,000 cards.
- 51-2 Math Teaching in Junior High Schools (quinquennial): Enrollment, facilities, teachers, contents of math programs. Reporting unit: public junior high schools. Unrestricted; not kept; 1962; 7,800 cards.
- 51-3 Science Teaching in Junior High Schools (quinquennial): Enrollment, science facilities, number of teachers. Reporting unit: public junior high schools. Unrestricted; not kept; 1962; 6,000 cards.
- 51-4 National Inventory of School Facilities and Personnel for Resource Evaluation and Damage Assessment (annual): Name, address, and location of schools, permanent buildings used for instructional purposes, number of permanent general use facilities, number of pupils and school employees. Reporting unit: public elementary and secondary school plants. Unrestricted; not kept; 1962; 12 reels.
- 51-5 Survey of Engineering Degrees (annual): By sex, engineering degrees according to curriculum and level. Reporting unit: institutions granting engineering degrees. Unrestricted; 2 years; 1961-62; 11,000 cards.
- 51-6 Survey of Engineering Enrollment (annual): For each of 26 curriculums in engineering, the number of undergraduate and graduate students by sex, year of study, and number of day and evening students. Reporting unit: institutions granting engineering degrees. Unrestricted; 2 years; 1962-63; 12,000 cards.
- 51-7 Offerings and Enrollments in NonPublic Secondary Schools (selected intervals): Enrollment, teaching staff and curriculum. Reporting unit: all nonpublic secondary schools. Unrestricted; not kept; 1961-62; 47,000 cards.
- 51-8 Offerings and Enrollments in Science and Mathematics in Public High Schools (biennial): Math and science courses and enrollment by sex. Reporting unit: public high schools. Unrestricted; 2 years; 1962-63; 9,000 cards.
- 51-9 Current Expenditures Per Pupil in Public Schools (annual): Per pupil expenditures for administration, instruction, attendance and health services, pupil transportation services, operation of plant, maintenance of plant, and fixed charges. Reporting unit: public school system. Unrestricted; 1 year; 1961-62; 3,000 cards.



- 51-10 Survey of Opening Fall Enrollment (annual): Full time, part time, and first time fall enrollment, by sex. Reporting unit: all institutions of higher learning. Unrestricted; 2 years; 1962-63; 12,000 cards.
- 51-11 Comprehensive Report on Enrollment (biennial): Number of resident and extension students, and students taking courses by TV or radio and enrolled in short courses or individual lessons. Also enrollments by sex in the preceding summer session. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1959-61; 55,000 cards.
- 51-12 Residence and Migration of College Students (quinquennial): Undergraduate and graduate students (full time and part time, and whether first time, new transfer, or continuing students) from each State or U.S. territory. Reporting unit: institutions with students taking creditable toward a bachelor's or higher degree. Unrestricted; 5 years; 1963; 100,000 cards.
- 51-13 Earned Degrees Granted During Year (annual): Four-year bachelor's and first professional degrees; first professional degrees requiring 5 or more years; second level degrees; and doctorates. Reporting unit: institutions granting bachelor's or higher degrees. Unrestricted; 2 years; 1962-63; 50,000 cards.
- 51-14 Faculty and Other Professional Staff (biennial): By sex, staff for general administration, student personnel services, resident instruction, extension instruction, organized research, elementary or secondary instruction and other facts. Reporting unit: all institutions of higher education. Unrestricted; 4 years; 1961-63; 24,000 cards.
- 51-15 Survey of Students Enrolled for Advanced Degrees (annual): Graduate students, both full and part time, enrolled for master degrees and doctorates by field of study and year level. Reporting unit: institutions granting advanced degrees. Unrestricted; 2 years; 1962-63; 20,000 cards.
- 51-16 Higher Education Planning and Management Data (annual): Faculty and administrative salaries, new buildings completed, and basic student charges. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1962-63; 47,000 cards.
- 51-17 College and University Library Statistics (annual): Library collections, personnel and expenditures. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1960-62; 20,000 cards.
- 51-18 Survey of Non-Public Elementary Schools (quinquennial): School description, enrollment by grade and number of secular and non-secular teachers. Reporting unit: nonpublic elementary schools. Unrestricted; 5 years; 1961-62; 14,500 cards.
- 51-19 Survey of Non-Public Secondary Schools (quinquennial): School description, enrollment by grade and number of secular and non-secular teachers. Reporting unit: nonpublic secondary schools. Unrestricted; 5 years; 1960-61; 14,000 cards.
- 51-20 Special Education for Exceptional Children (quinquennial): Enrollment and grade level of blind, partially seeing, hard of hearing, speech impaired, crippled, special health problems, socially and emotionally maladjusted, mentally retarded, and gifted students. Also, number of teachers. Reporting unit: public school systems and residential schools. Unrestricted; 5 years; 1962-63; 67,000 cards.
- 51-21 Inventory of College and University Facilities (quinquennial): Existing physical facilities of institutions of higher education. Reporting unit: institutions of higher education. Unrestricted; 5 years; 1957; 150,000 cards.
- 51-22 Public School Library Statistics (biennial): Library collections, number of school librarians, expenditures for books and number of school libraries. Reporting unit: public school systems. Unrestricted; 4 years; 1960-62; 75,000 cards.
- 51-23 Public Library Statistics (quinquennial): Library collections, personnel and expenditures. Reporting unit: all public libraries. Unrestricted; 5 years; 1962; 68,000 cards.

- 51-24 Financial Statistics of Institutions of Higher Education (biennial): Amounts and sources of income, expenditures by purpose, property utilized, endowment funds, students' loan funds, and annuity and living trust funds. Reporting unit: higher education institutions. Unrestricted; 4 years; 1959-61; 140,000 cards.
- 51-25 Organized Occupational Curriculums (biennial): Enrollments and graduates of organized occupational curriculums. Reporting unit: institutions of higher education. Unrestricted; 4 years; 1959-61; 50,000 cards.
- 51-26 Education Directory, Part 2 (annual): Total enrollment, superintendent's name and address. Reporting unit: public school systems having an enrollment of 300 or more. Unrestricted; 1 year; 1963; 28,000 cards.
- 51-27 Offerings and Enrollments in High School Subjects (decennial): Enrollment by subject and grade level. Reporting unit: public secondary schools. Unrestricted; 10 years; 1960-61; 240,000 cards.
- 51-28 Statistics of College and University Libraries (biennial): Collection, staff, expenditures and individual salary by level of personnel. Reporting unit: heads of academic libraries. Confidential; 2 years; 1962-63; 11,500 cards.
- 51-29 Public School Library Statistics (quinquennial): Resources, staff, space, level of service, enrollment. Reporting unit: heads of school library administrations. Confidential; 2 years; 1960-61; 60,000 cards.
- 51-30 Statistics of Public School Libraries (biennially): Resources, enrollment, staff, expenditures. Reporting unit: heads of school library administrations. Confidential; 2 years; 1962-63; 3,400 cards.
- 51-31 Statistics of Public Libraries (quinquennial): Resources, staff, expenditure, and service area. Reporting unit: heads of public libraries. Confidential; 2 years; 1962; 79,450 cards.
- 51-32 Statistics of Local Public School Systems (one-time survey): Instructional personnel, public facilities, and fiscal data. Reporting unit: public school systems. Unrestricted; 2 years; 1959-60; 138,600 cards.
- 51-33 Survey of Home Economics (biennial): Administration and organization of home economics programs, enrollment and faculty. Reporting unit: 4-year institutions which offer programs in home economic leading to at least a bachelor's degree. Unrestricted; 2 years; 1963; 3,550 cards.
- 51-34 Status and Career Orientations of College Faculty Members (One-time survey): Personal characteristics, position, and assignment, educational background, work experience, economic status and occupational plans of college faculties. Reporting unit: higher education institutions. Unrestricted; 2 years; 1962-63; 32,000 cards.
- 51-35 Organization and Administration of Student Personnel Services (one-time survey): Background of selected student services personnel, staffing of student services area, pattern of organization and manner in which policies are approved and implemented. Reporting unit: institutions of higher education. Unrestricted; 2 years; 7,800 cards.
- 51-36 Science Teaching in Elementary Schools (one-time survey): Objectives, enrollment by grade, facilities, teaching aids, practices and procedures. Reporting unit: elementary schools. Unrestricted; 2 years; 1960-61; 43,500 cards.
- 51-37 Status of Industrial arts in Public Secondary Schools (one-time survey): Objectives, classes, enrollment by sex, laboratories, teachers, methods and problems within the curriculum area of industrial arts. Reporting unit: public secondary schools. Unrestricted; 2 years; 1962-63; 9,150 cards.
- 51-38 Survey of Early Elementary Education in Public Schools (one-time survey): Status, characteristics, practices, and policies of early elementary education; i.e., nursery, kindergarten, and primary grades. Reporting unit: medium-sized and larger public school systems. Unrestricted; 2 years; 1960-61; 1 tape.

- 51-39 Foreign Languages in Public Secondary Schools (one-time survey): Courses taught, enrollment, materials, equipment, educational background of teachers and salaries. Reporting unit: public secondary schools. Unrestricted; 2 years; 1959; 43,000 cards.
- 51-40 Fifty-Year Programs: A survey of Policy and Practice in Classroom-Teacher Education (one-time survey): Policies and practices in programs at the fifth-year level designed for elementary and secondary classroom teacher education. Reporting unit: higher education institutions having fifth-year teacher training programs. Unrestricted. 2 years; 1959-60; 19,000 cards.

## BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

- 55-1 Survey of Financial Characteristics (one-time survey): Components of net worth. Reporting unit: Family units. Confidential; 3 months; 1962; 33 tapes.
- 55-2 Federal Funds Historical Data (daily and weekly): Basic reserve position, Federal funds and related dealer transactions. Reporting unit: individual member banks. Confidential; permanent; 1959-present; 6 tapes.
- 55-3 Cost of Living for Latin America and Western Europe (annual): Index of cost of living for approximately 25 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 250 cards.
- 55-4 Exports of Goods and Services, Growth Rates for Asia, Latin America, and Western Europe (annual): Single total exports of goods and services for approximately 50 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 600 cards.
- 55-5 World Exports, Compound Annual Growth Rates (annual): Single total of exports for approximately 112 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 1,120 cards.
- 55-6 World Imports, Compound Annual Growth Rate (annual): Single total of imports for approximately 112 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1950-63; 1,456 cards.
- 55-7 Real Gross National Product, Growth Rates for Asia, Latin America, and Western Europe (annual): Real GNP for approximately 50 countries. Reporting unit: U.N. Yearbook of National Accounts Statistics and official country data. Unrestricted; permanent; 1950-62; 600 cards.
- 55-8 Gold and Foreign Exchange Reserves, Growth Rates for Latin America and Western Europe (annual): Single total of reserves for approximately 40 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 400 cards.
- 55-9 Industrial Production, Growth Rates for Latin America and Western Europe (annual): Composite index of industrial production for approximately 15 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 150 cards.
- 55-10 Money Supply, Growth Rates for Latin America and Western Europe (annual): Single total of money supply for approximately 25 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 250 cards.
- 55-11 U.S. Exports to Asia (monthly): Single total exports to Asian area (Pakistan east through Japan). Reporting unit: Bureau of the Census, U.S. Department of Commerce. Unrestricted; permanent; 1953-64; 72 cards.
- 55-12 Japanese Industrial Activity Index (monthly): Composite index of output in manufacturing, mining, and utility sectors. Reporting unit: U.S. Embassy, Tokyo, Japan. Unrestricted; permanent; 1947-63; 34 cards.
- 55-13 Savings Flows to Savings and Loan Associations and Mutual Savings Banks (annual): Levels, new deposits, withdrawals of savings capital at savings and loan associations and regular deposits at mutual savings banks. Reporting unit: Federal Savings & Loan Insurance Corp. & National Association of Mutual Savings Banks. Unrestricted; permanent; 1964; 1,500 cards.

- 55-14 Boeckh Index of Construction Costs (monthly): Residences and apartments—building costs. Reporting unit: builders. Unrestricted; permanent; 1934-59; 600 cards.
- 55-15 Wholesale Price Indexes—Construction Materials (monthly): Construction materials and components. Reporting unit: wholesalers. Unrestricted; permanent; 1947-59; 6,000 cards.
- 55-16 Mortgage Recordings (monthly): Recordings by type of holder and type of mortgage. Reporting unit: Mortgage Recording Offices and other sources available to Home Loan Bank Board. Confidential; permanent; 1939-59; 4,000 cards.
- 55-17 FHA—Insured Home-Mortgage Terminations (monthly): Date of origination, data of termination, characteristics of borrower, etc. Reporting unit: FHA-insurance offices. Confidential; permanent; 1956-62; 1,000,000 cards; 25 tapes.
- 55-18 Mortgage Debt—Net Changes Only (quarterly): One-four family mortgage debt; multifamily and commercial mortgage debt; farm mortgage debt. Reporting unit: financial institutions and other mortgage holders. Confidential; permanent; 1949-63; 500 cards.
- 55-19 Reports of Condition, All Insured Commercial Banks (quarterly): Asset and liability items—in detail. Reporting unit: individual banks. Confidential; permanent; 1959-63; 580,000 cards; 29 tapes.
- 55-20 Sample Survey of Agriculture (one-time survey): Selected data on farm debts and other characteristics of the farm and farm operators. Reporting unit: farm operators and landlords in the United States. Confidential; 5 years; 1960; 264,000 cards; 2 tapes.
- 55-21 Reports of Income and Dividends. All Insured Banks (annual): Breakdown of income, expenses, taxes, etc. Reporting unit: individual banks. Confidential; permanent; 1960-63; 240,000 cards; 8 tapes.
- 55-22 Operating Ratios—Member Banks (annual): Forty ratios showing relationship between various balance sheet items and income and expense items. Reporting unit: individual banks. Confidential; 2 years; 1962-63; 18,000 cards.
- 55-23 Electric Power Series (monthly): KWH's. Reporting unit: electric utilities and industrial self-generators. Confidential; permanent; 1957 to present.
- 55-24 Industrial Generation of Electricity (monthly, annual): KWH's—current generation capacity. Reporting unit: Federal Power Commission. Unrestricted; permanent, 1957 to present; 100,000 cards; 1 tape.
- 55-25 Federal Funds Rate (daily): Effective rate, low bid, high offer. Unrestricted; 10 years; 1955-62; 1,800 cards.
- 55-26 U.S. Government Securities Dealer Statistics (daily): Positions, borrowings and transactions. Confidential; permanent; 1963 to present; 40,000 cards; 2 tapes.
- 55-27 U.S. Government Security Yields and Prices (monthly): Yields on key Treasury bills and average yields and prices on Treasury coupon issues. Unrestricted; 10 years; 1952-63; 3,000 cards.
- 55-28 Statement Week Averages—3 Month Bill Rate and Reserves (weekly): Free reserves; bill rate; weeks high, low; spread; excess reserves; borrowed reserves; 3 weekly moving average; free reserves; 9-week moving average free reserves; spread as percent of bill rate; deviation from 3-week and 9-week average; 8- and 13-week lag in moving averages; net changes in variables. Confidential; 10 years; 1953-63; 2,500 cards.
- 55-31 Money Supply (daily): Due to and from banks, Federal Government deposits, other demand, vault cash, cash items, time deposits. Reporting unit: individual Reserve banks. Unrestricted; permanent; 1958 to present; one tape.
- 55-32 Survey of Negotiable Time Certificates of Deposit (one-time survey): Outstanding certificates of deposit with breakdowns by denominations, original maturity, and type of holder. Reporting unit: individual member banks and selected nonmember banks. Confidential; permanent; 1960-62; 600 cards.

- 55-33 Member Bank Deposits and Vault Cash (daily): Demand deposits: due to banks, U.S. Government demand, other demand, net demand; time deposits; vault cash. Reporting unit: Federal Reserve District summaries. Confidential; permanent; 1958 to present; 58,000 cards.
- 55-34 Flow-of-Funds Accounts (quarterly): Elements of the published accounts. Reporting unit: Flow-of-Funds Section, Division of Research and Statistics. Unrestricted; 1 year; 1946-63; one tape.
- 55-35 Spot Exchange Rates: Major Currencies Against U.S. Dollar (weekly): Quotations on Swiss, German, I.K., Dutch, French, Italian, Canadian, Belgian and Japanese currencies. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-36 Industrial Stock Indices (weekly): Indexes for industrial stock prices in Canada, United States, United Kingdom, Japan, Germany and Switzerland. Reporting unit: market data. Unrestricted; permanent; 1958 to present; 352 cards.
- 55-37 Long-Term Bond Yields (weekly): Yields on specific long-term bonds in United Kingdom, United States, Germany Canada and Switzerland. Reporting unit: market data. Unrestricted; permanent; 1958 to present; 320 cards.
- 55-38 Short-Term Interest Rates (weekly): Yields on German, United Kingdom, United States, and Canadian Treasury bills; Swiss 3-month deposits; and Japanese bank loans and discounts. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-39 Interest Arbitrage for German Commercial Banks (quarterly): Yields on Euro-Dollar Deposits (London), German inter-bank loan rate, German Treasury bills, forward rates on DM, and differences. Reporting unit: market data. Unrestricted; permanent; 1960 to present; 214 cards.
- 55-40 Interest Arbitrage, New York/London (weekly): Yields on United States and Canadian Treasury bills, forward rates on Canadian dollar, and difference. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-41 Interest Arbitrage New York/Canada (weekly): Yields on United States and United Kingdom Treasury bills, United Kingdom local authority deposits, forward rates on sterling, and differences. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-42 International Money Market Yields for U.S. Dollar Investors (weekly): Yields on Euro-\$ deposits (London); United States CD's; Canadian, United States and United Kingdom Treasury bills; and United Kingdom hire purchase, Canadian Finance Co. and U.S. Finance Co. paper. Reporting unit: market data. Unrestricted; permanent; 1963 to present; 85 cards.
- 55-43 Three-Month Forward Exchange Rates (weekly): Forward quotations on Swiss, German and United Kingdom currencies against the United States dollar and Swiss, United States, Germany, Dutch, Belgian and French currencies against sterling. Reporting unit: market data. Unrestricted; permanent; 1961 to present; 160 cards.
- 55-47 F.R. Bulletin Profit Series (quarterly, annual): Sales, profits before taxes, taxes, profits after taxes, dividends. Reporting unit: 180 manufacturing corporations. Unrestricted; permanent; 1955-59; 4,500 cards.
- 55-48 Statistical Data Compiled From Bank Examination Reports (quarterly): Principal assets and liabilities; maximum interest rates on time deposits; maturity distribution of investments; amounts of criticized assets and gross losses on loans; percentage of past due loans and other ratios. Reporting unit: individual member banks (one report per year for each bank). Confidential; permanent; 1963-Present; 68,00 cards; 5 tapes.
- 55-49 Member Bank Loans to Commercial and Industrial Borrowers: One time survey. Amount and term of loan and borrower characteristics of each loan. Reporting unit: individual bank data reported by a sample of about 2,000 banks. Confidential; permanent; 1955; 4 tapes.

- 55-50 Member Bank Loans to Commercial and Industrial Borrowers: One time survey. Amount and terms of loan and borrower characteristics for each loan. Reporting unit: individual bank data reported by a sample of about 2,000 banks; confidential; permanent; 1957; 4 tapes.
- 55-51 Interest Rates Charged by Banks on Short-Term Business Loans (quarterly): Amount of loan, annual interest paid and annual rate of interest for new or renewal loans. Reporting unit: individual bank data reported by sample of banks in 19 leading cities; confidential; permanent; 1958-63; 3 tapes.
- 55-52 Interest Rates Charged by Banks on Short-Term Business Loans (quarterly): Amount of loan, annual interest paid and annual rate of interest for new or renewal loans. Reporting unit: individual bank data reported by sample of banks in 19 leading cities; confidential; permanent; 1964; 2 tapes.
- 55-53 Ownership of Demand Deposits Survey (annual): Number of accounts and amount of demand deposits held by individuals, partnerships and corporations grouped by major types of holders and account size. Reporting unit: individual bank data from a sample of about 1,800 banks; confidential; permanent; 1959-61; 6 tapes.

## DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

## PUBLIC HEALTH SERVICE, NATIONAL CENTER FOR HEALTH STATISTICS

- 68-1 National Lung Cancer Mortality Survey (one time survey): Data include smoking history, residence history (identification of places by population size in which the deceased person resided 10 years or longer), diagnostic techniques used to establish diagnosis, and histologic type of cancer. Reporting unit: informants listed on death certificates, including family informant, attendant physician and institution in which death occurred. Data collected for a sample of approximately 3,000 lung-cancer deaths. Confidential; permanent; 1958; 4,000 cards.
- 68-2 National Mortality Survey (one time survey): Data include items analogous to those collected in the 100-percent and 25-percent stages of the census of population, including place of residence of decedent, family type and family status of deceased person, related information about decedent's family, employment status, occupation, and industry of deceased person and his spouse, if married, income of deceased, of his spouse, and other family members. Reporting unit: family informants listed on death certificates. Data collected for a sample of approximately 10,000 deaths. Confidential; permanent; 1960; 2 tapes.
- 68-3 National Mortality Sample Survey (annual): Hospital use during last year of life; diagnostic information and information on operations for each episode of medical care; varying from year to year, social, economic, and family characteristics of deceased persons, such as family income, educational attainment of deceased person and spouse, place of previous residence. Reporting unit: informants listed on death certificates; records of hospitals and resident medical care institutions. Confidential; permanent; 1961 to present; 4 tapes.
- 68-4 National Natality Sample Survey (annual): Survey topic varies from year to year. Topics covered to date include radiation exposure during pregnancy, medical care during pregnancy, family status of mother, socioeconomic characteristics of mother, and expectation of additional births. Reporting unit: Informants listed on birth certificates, including mother, hospital of birth, attendant physicians. Confidential; permanent; 1963 to present; 2 tapes.
- 68-5 Master Facility Inventory (annual): Name, address of establishment, ownership, type of service, number of beds, and number of employees. Reporting unit: hospital or resident institutions. Confidential; permanent; 1962 to present; 5 tapes.
- 68-6 Resident Places Survey I (one time survey): Admission policy, utilization statistics, and health of residents. Reporting unit: nursing and personal care homes and residents of these homes. Confidential; permanent; 1962; 2 tapes.

- 68-7 Health Interview Survey (annual): Persons, acute and chronic conditions, hospitalization experience, other health characteristics. Reporting unit: household members in civilian, noninstitutional population. Confidential; permanent; 1957-64; 35 tapes.
- 68-8 Health Interview Survey—Hearing Ability Supplement (one-time survey): Hearing ability, history of hearing problem; use of hearing aid, lip reading, sign language. Reporting unit: household members with hearing impairment in civilian, noninstitutional population. Confidential; permanent; 1962-63; 2 tapes.
- 68-9 Health Interview Survey—Personal Health Expenditure Supplement (one-time survey): Expenditures by persons in household for doctors' bills, hospital bills, medicines, dentists' bills, and other medical expenses. Reporting unit: household members in civilian, noninstitutional population. Confidential; permanent; 1962; 2 tapes.
- 68-10 Health Examination Survey Data From First Cycle Program (one-time survey): Findings of individual health examinations, and related household questionnaires and control records. Reporting unit: individuals selected to constitute probability sample of U.S. adults, ages 18-79. Confidential; 1959-62; 300,000 cards; 25 tapes.
- 68-11 Health Examination Survey Data From Second Cycle Program (one-time survey): Findings of individual health examinations and related household questionnaires and control records. Reporting unit: individuals selected to constitute probability sample of U.S. children, ages 6-11. Confidential; 1963-65; 400,000 cards.
- 68-12 Birth and Fetal Death Statistics (annual): County, city, sex, race, birth order, ages of parents, month, birthweight, legitimacy, and other information. Reporting unit: State, territorial, and independent city health departments. Confidential; 2 years; 1960-63; 8,000,000 cards; 87 tapes.
- 68-13 Marriage Statistics (annual): County, race, age, previous marital status, number of marriages for each party: month. Reporting unit: State, territorial, and independent city health departments. Confidential; 1960-62; 127,000 cards; 3 tapes.
- 68-14 Death Statistics (annual): County, city, sex, race, age, month, nativity, marital status, cause of death. Reporting units: State, territorial, and independent city health departments. Confidential; 5 years, 1950-63; 8,700,000 cards; 70 tapes.
- 68-15 Divorce Statistics (annual): County, month, duration of marriage, legal cause, race, age, previous marriage, number of children. Reporting unit: State, territorial, and independent city health departments. Confidential; 1960-62; 54,000 cards.
- 68-16 Birth, Death, and Fetal Death Statistics (annual): Sex, color, age, birth order, cause of death. Reporting unit: generated in data-processing operations. Unrestricted; permanent; 1947-59; 4,160,000 cards.
- 68-17 Special Death Statistics by Occupations and Industry (one-time survey): State, occupation, industry, age, color, cause of death (all males 20-64 years). Reporting unit: State, territorial, and independent city health departments. Confidential; permanent; 1950; 335,000 cards.
- 68-18 Special Death Statistics by Occupations and Industry (one-time survey): State, sex, color, age, up to five causes of death. Reporting unit: State, territorial, and independent city health departments. Confidential; permanent; 1955; 500,000 cards.

## SOCIAL SECURITY ADMINISTRATION

- 72-1 One-Percent Sample Employee-Employer Record (annual): Year of birth, sex, race, industry and geographic codes, quarterly and annual wages. Reporting unit: social security account number holder. Confidential; 10 years; 1955-62; 90 tapes.

- 72-2 One-Percent Continuous Work History Sample (annual): Year of birth, sex, race, earnings each year, benefit and insurance status, total earnings and coverage since 1936. Reporting unit: social security account number holder. Confidential; 3 years; 240 tapes.
- 72-3 One-tenth of One Percent Continuous Work History Sample (annual): Year of birth, sex, race, earnings each year 1937 to date, benefit and insured status. Reporting unit: social security account number holder. Confidential; 3 years; 36 tapes.
- 72-4 Name and Address File Tape (quarterly): Name and mailing address of employers. Reporting unit: employers reporting wages. Confidential; permanent; 70 tapes.
- 72-5 Data Employer Duplicate Check Tape (quarterly): Wages and wage items reported by employers for 5 quarters. Reporting unit: employer. Confidential; permanent; 165 tapes.
- 72-6 Summary Earnings Record Tape (quarterly): Annual earnings 1951 to date, total earnings 1937 to date, quarters of coverage 1951 to date, date of birth, sex. Reporting unit: social security account number holder. Confidential; permanent; 1,500 tapes.
- 72-7 Regular Transcript Master Benefit Tape (monthly): Benefit amount, age, State and county of payee, technical and historical data on benefits. Reporting unit: person entitled to OASDI monthly benefit. Confidential; permanent; 400 tapes.
- 72-8 Employer Identification Tape (semiannual): Geographic and standard industrial classification (SIC) codes. Reporting unit: county-industry reporting units for each employer. Confidential; permanent; 34 tapes.

## APPENDIX 2.—STATISTICAL EVALUATION REPORT NO. 6—REVIEW OF PROPOSAL FOR A NATIONAL DATA CENTER

(A Report Prepared by Edgar S. Dunn, Jr., Consultant to the Office of Statistical Standards Bureau of the Budget)

### PREFACE

This report "Review of Proposal for a National Data Center" is the sixth of a series presenting the results of a comprehensive review and evaluation of some aspects of the statistics program of the Federal Government. It was prepared by Edgar S. Dunn, Jr., Resources for the Future, Inc., as consultant to the Bureau of the Budget.

The proposal which Mr. Dunn has reviewed stems from the work of a committee established by the Social Science Research Council to study the problems of the preservation and use of economic data. In the spring of 1965 that committee made its report to the SSRC, which presented it to the Director of the Bureau of the Budget. In its report, the Committee on the Preservation and Use of Economic Data, known as the Ruggles committee, summarized its recommendations as follows: "First, \* \* \* that the Bureau of the Budget, in view of its responsibility for the Federal statistics program, immediately take steps to establish a Federal Data Center. \* \* \* Second, that the Office of Statistical Standards \* \* \* place increased emphasis on the systematic preservation in usable form of important data prepared by those agencies engaging in statistical programs. \* \* \* Third, that at an early date the Social Science Research Council convene representatives from research institutions and universities in order to develop an organization which can provide a clearinghouse and coordination of requests for data made by individual scholars from Federal agencies."

In asking Mr. Dunn to examine the proposal, and to study ways of implementing it, we were concerned primarily with the first two of these recommendations. But, while the Ruggles committee represented the interests of the academic and social science research community, we were concerned with the use of statistical data for research, policy and decisionmaking at all levels, both within and outside Government. Mr. Dunn wisely extended it to include consideration of the relationships between the collecting and compiling processes on the one hand and preservation and accessibility for further use on the other hand.

We are indebted to Mr. Dunn for this analysis and report. We are also indebted to the many persons with whom he consulted, particularly those mentioned in his letter of transmittal who assisted by their thoughtful analysis of particular aspects of the entire problem and the preparation of the important appendix material.

RAYMOND T. BOWMAN,  
*Assistant Director for Statistical Standards.*

NOVEMBER 1, 1965.

Dr. R. T. BOWMAN,  
*Assistant Director for Statistical Standards,  
U.S. Bureau of the Budget,  
Executive Office Building, Washington, D.C.*

DEAR RAY: Transmitted herewith is the final report containing my review of the proposal for a National Data Center.

In seeking to identify ways of implementing this proposal I undertook an intensive period of study and review covering the last 6 months. I had the benefit in this effort of the advice and counsel of numerous others. This assistance was engaged in the following way. I divided the problem into parts that could be considered simultaneously. One part consisted of an informal ad hoc committee which met with me on a number of occasions to discuss the feasibility of establishing a referral-reference function in relation to the files of the Federal

statistical system. Joining me in this discussion were Joe Daly and Ed Goldfield of the Bureau of the Census, Bob Steffes and Rudy Mendelsohn of BLS and Ezra Glaser of the Patent Office. In a second initiative Mendelsohn undertook a more intensive study based upon the earlier survey of machine readable records conducted by OSS to try to get a better fix on what it will take to establish an archival function. He in turn was assisted by many people in the agencies. Lastly, the National Bureau of Standards was used as a vehicle to assemble a small group of knowledgeable people in an attempt to specify more clearly the essential elements of a data service center required to provide a range of facilitating services. Ezra Glaser, Marshall Wood, and Dave Rosenblatt were the principal contributors to this effort although conversations included Sam Alexander and other members of his staff. Paul Krueger and I also participated in this effort.

In addition, I have engaged in many discussions of substantive issues with a number of knowledgeable people in the Federal agencies (both statistical agencies and program agencies) and in the universities.

I particularly want to acknowledge the invaluable assistance that I have received from Paul Krueger on your staff. He has given me continuous support, assistance and encouragement. He has joined me in many of the meetings and discussions with the aforementioned and has made his own valuable contribution to the thinking process.

The form and content of the report, of course, remains my own responsibility. I believe that the general conclusions and recommendations are sound and supported in whole or in part by the informed judgment of many others beside myself, but I do not attribute the views of this report specifically to any one or all of its many contributors.

In writing the report I have incorporated material included in earlier memorandums as well as sections that contain explanation and argument that is superfluous from the point of view of the informed staff member of the Office of Statistical Standards. I did so because I assumed that this report might be used in whole or in part to communicate elements of this problem and the recommended solutions to more than one group. I attempted, therefore, to include a comprehensive discussion of the problems and opportunities.

The report makes clear that my own understanding and evaluation of this problem has modified somewhat in the course of the study. I now feel that the production standards and practices are a more important element in both problems and solution. This, as well as other considerations, leads me to be less sanguine about the possibility or the desirability of keeping the issues of organization in the background. I think that there might be some benefit in our discussing this and several related issues on an informal basis.

Let me say that I have enjoyed working with the Office of Statistical Standards on this problem. I hope that the results are constructive in serving your needs and objectives.

Sincerely yours,

EDGAR S. DUNN, JR.

### SUMMARY

The Ruggles Committee report recommending the establishment of a National Data Center is only one of the more manifest expressions of concern, dissatisfaction and frustration that have been surfacing among the groups that use numerical records for research, planning or decisionmaking at all levels. The problems at issue go far beyond the forms of discontent generated by special interests or marginal interests not served by public policy. They result from major changes on both the demand and supply side of the information process since World War II. Many people in this wider circle are attached to a rather naive data bank concept of the solution that does not incorporate an adequate appreciation of the basic problems in data use and data generation.

### THE PROBLEM

The central problem of data use is one of associating numerical records and the greatest deficiency of the existing Federal statistical system is its failure to provide access to data in a way that permits the association of the elements of data sets in order to identify and measure the interrelationship among interdependent or related observations. This is true at virtually all levels of use and for all purposes from academic model builders to business market researchers.



There are a number of characteristics of existing programs and procedures that stand in the way of an effective association of numerical records for purpose of analysis.

(1) Important historical records may be lost because of the absence of consistent policy and procedure for establishing and maintaining archives.

(2) The absence of appropriate standards and procedures for file maintenance and documentation lead to low-quality files that contain many technical limitations to effective association of records.

(3) Many of the most useful records are produced as a byproduct of administrative or regulatory procedures by agencies that do not recognize a general purpose statistical service function as an important part of their mission.

(4) Record association requires a good deal of intelligence about the compatibility of records in several dimensions and the circumstances that condition their availability. There exists no organized reference capability for performing the kind of reference service essential for the Federal system or even within individual agencies.

(5) Production procedures for collecting, coding and tabulating data that were appropriate when developed now lead to several types of record incompatibility that block the kinds of record association in usage that is required by current policy problems and made possible by computer technique.

(6) There are identifiable gaps in existing data records that stand in the way of bringing together records of greatest relevance for today's problems. Some of these gaps are more apparent than real and reflect the effect of the other obstacles to effective record association.

(7) The structural problems of concern to today's policymakers and the effort to bypass problems of record incompatibility force the utilization of data at levels of disaggregation that place severe strains upon regulations restricting the disclosure of information about individual respondents. Technical possibilities for using the computer to bypass these disclosure constraints have not been generally developed and made available by the agencies.

(8) There are new possibilities for more efficient management of large-scale numerical files in terms of storage and retrieval; new possibilities for rearranging files in more useful form; new possibilities for retrieving in the form of maps, graphs, charts and other media in addition to the traditional tabular forms; new possibilities for building in disclosure controls and disclosure bypasses; new possibilities for matching records to assure compatibility. These potentialities require the expenditure of time and effort on system design and software development that few agencies can justify.

#### THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability and compatibility are very high. This rests in part upon an unexploited joint demand for information and an information service capability. In the Federal domain alone large amounts of money are being ineffectively spent in an effort to deal with these problems. The amount of overlapping and resource waste is substantial. The stakes are also high because the improvement in the utility of the information base could have an unmeasurable but substantial effect upon the quality of public administration.

#### THE SOLUTIONS ARE NOT SIMPLE

The solution to these problems do not rest, as some think, in bringing a large number of tapes into a common repository. Nor does it rest upon the fact that many different uses impact upon the same data sets making them "general purpose" records susceptible to central management. General-purpose files are always put to special-purpose uses. What makes a record a general-purpose record is for it to be constructed on the basis of standards, maintained in effective condition and serviced by institutional arrangements and a technical system capability that will allow it to be combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users. Thus the solution to the problems will require program modification on a broad front involving all of the agencies as well as an emergent data service center.

#### RECOMMENDATIONS

Accordingly, it is recommended that a National Data Service Center be established with the capability to:

- (1) Manage archival records;
- (2) Develop referral and reference services;
- (3) Provide explicit facilitating services for users including:
  - (a) File rearrangement, cost tabulation and extended output options;
  - (b) Tape translation and file modification;
  - (c) Record matching;
  - (d) Disclosure bypassing; and
  - (e) Standard statistical routines.
- (4) Develop computer hardware and software systems essential to above;
- (5) Provide staff support to work in conjunction with the Bureau of the Budget to develop and establish and monitor standards essential to the system capability; and
- (6) Establish a research capability directed to an analytical evaluation of user requirements for the purpose of designing and developing the system components essential to perform these services.

The National Data Service Center would perform these services for:

- (1) Archival records under direct management control of the Center;
- (2) The current and accumulated records of administrative and regulatory agencies;
- (3) As a system resource to be used in connection with the current records of any agency not in a position to meet the needs independently.

#### RESOURCE REQUIREMENTS

Figures are offered that represent estimates of program costs for several components of these functions but it is pointed out that there are joint-product, joint-cost relationships between these service activities that make these estimates questionable as a guide to overall program costs. A judgment is made that the range of services and program adjustments required (including resources for modifying agency programs to be consistent with system requirements) would call for expenditure of between \$1 and \$2 million annually during the first year or two and rising to the neighborhood of \$10 million annually over a period of 5-10 years. A serious problem will be the assembly of the kind of intellectual resources required.

#### ISSUES OF ORGANIZATION

All of this really raises the issue of what kind of Federal statistical system we want to develop in the next generation, and encompasses a number of issues of organization and mission that will need to be addressed in a broad context. Further progress on the whole effort must depend on some understanding of the issues at the top policy level and some preliminary policy decisions to guide the direction of further effort.

#### PRIORITIES

If an effort is made to undertake this kind of system development there are certain immediate requirements that need to be fulfilled.

- (a) A continuous focus of leadership needs to be established.
- (b) This focus needs to be provided with staff support providing the kind of research-analytic capacity that can evaluate use requirements as a guide to specifying program options and reduce these options to specifications, costs and a logical order of time phasing.
- (c) Begin under existing authority to:
  - (1) Develop the standards for archives and compatible statistical building blocks, and
  - (2) Begin work on the 9,000 tape nucleus archive identified in appendix B. The Bureau of the Budget should give serious consideration to requesting funds to support these efforts in the fiscal 1967 budget.

#### INTRODUCTION

The assignment leading to this report originated 6 months ago in the form of a request to seek out and identify ways of implementing the proposal for a National Data Center presented by the Subcommittee on the Preservation and Use of Economic Data of the Social Science Research Council (the Ruggles

committee). During the interim the proposal and the problem set to which it was addressed has been intensively reviewed. I have had the benefit in this effort of the Council and effort of numerous others as indicated in the covering letter.

In order to identify the program options that might satisfy the intent of the Ruggles committee proposal, it was necessary to try to specify more precisely the problem set at issue. This has led to a characterization of the problem and attendant proposals in a somewhat broader context than the committee report but in a way that seems consistent with the intent and the leadership council of that group.

It is important to note that the report of the Ruggles committee is only one of the more manifest expressions of concern, dissatisfaction and frustration that have been surfacing among the groups that use numerical records for research, planning, or decisionmaking at all levels. The committee report is a good representation of the interests and the concern of the academic social science research community. There are other loci of discontent. The newly emerging welfare agencies of the Federal Government (OEO, EDA, the Department of Urban Affairs, the new programs in education and health in HEW, etc.) are experiencing great difficulty in assembling the statistical data that will guide them in analyzing their missions, establishing standards for performance and formulary for guidance, and support of administrative decisions and evaluating results. Their problems in this regard, when added to the usual difficulties of new program development, threaten to delay programs and render decisions more vulnerable to attack. A wide assortment of groups associated with making policy and planning for public facilities at all levels are becoming increasingly vocal about this concern. These are the groups that plan for roads, schools, hospitals, urban and regional development, etc. As business management turns increasingly to supplement its internal sources of information with the intelligence afforded by public agencies, they, too, are becoming aware of some of the inadequacies and anomalies of the information base.

Public needs for general purpose statistical information have never been satisfied and, indeed, never will. There are serious and legitimate issues of policy about how far down the scale one goes before general purpose becomes special purpose and about the levels of support for public information services that are appropriate. However, the problems that are at issue at the present moment go beyond the forms of discontent generated by marginal interests not served by current public policy. They are the product of fundamental changes on both the demand and supply side of the information process that have come to a head since World War II.

The most dramatic and obvious change on the supply side results from the advent of the large-scale computer. The economic feasibility and technical capability of producing, managing, and utilizing large numerical files has been multiplied by factors of a thousand in some technical applications and often by factors of 10 and a hundred in the economic and engineering dimensions of program planning. These are order of magnitude changes in capability that have come with revolutionary speed. They not only represent important and discontinuous changes in scale but also changes in kind because program options become technically feasible that were unthinkable as recently as 10 years ago.

Less commonly noted has been an increasingly dramatic change in information requirements on the demand side. In part, this is a consequence of the technical capabilities offered by the computer as well. The user can now handle data matrixes of a size and complexity formerly unmanageable and can use analytical techniques of a computational dimension formerly impossible. In part, however, the changes in information requirements stem from radical changes in demand factors distinct from these responses to expanded technical capability. Public policy in recent years has turned increasingly to a concern about the problems of social structure as they relate to public welfare and public policy. The issues of poverty, education, health, area depression, urban organization, etc., all require an increase in relevant detail for sub-system components of the total economy or total culture. At the same time the analytical disciplines in the social sciences and management analysis and control have been turning increasingly to quantitative methods and procedures.

As a result of these processes the users are increasingly finding that their needs for data are not satisfied by traditional documentary formats and the producers of data are finding the need to make data available more commonly in machine readable form—often in an organization and a format unique to the

purpose. The producers and users of data find their requirements and their missions intersecting in ways unknown a short time ago. Neither can continue to live the independent life formerly customary. The computer and other dimensions of social change have performed a shotgun wedding and both parties are in the process of discovering their incompatibility.

The procedural and program difficulties that led to this report are the product of this incompatibility. It is not uncommon in cases of this type for the assessment of difficulty to be one sided. The community of users has been sensitive to and vocal about many of the limitations of the producers of general purpose data—their mission concepts and institutional forms. The producers, in turn, can perceive many inadequacies on the part of the users. This report continues to be one sided in emphasis because it addresses itself primarily to the problems of use. The charge that framed it was couched in terms of the missions of the producing agencies of the Federal statistical system. It would be very worthwhile for some group to produce a companion evaluation of the anomalies in the production and usage of statistical information that arise from the practices and concepts of the user groups.

It is sometimes true that people who have concerned themselves with this problem are content with a superficial level of diagnosis and prescription. There is considerable attachment to the notion that most of our problems can be solved by computerizing all of the data we have in the backroom. This may be characterized as the "naive data bank" notion and its widespread acceptance is a source of some concern. I should emphasize here that this is not a characteristic of the Ruggles Committee report which was produced by knowledgeable and sophisticated people. However, the tendency to see the solution to the problems in relatively cheap technical programs has led to some misinterpretation of the Ruggles Committee report on the part of both those looking for additional support for data bank schemes and those reacting to the naive data bank concept. These evaluations and solutions are not based upon an adequate understanding and appreciation of the realities of the production processes essential to data generation or the institutional forms appropriate to their purpose.

The Ruggles Committee report gave us a healthy beginning toward an evaluation of this problem in realistic terms. However, this group did not have the time or staff resources to spell out the total problem set in a way that seems essential to support a more detailed consideration of program options. There is also a tendency in this report to see the problem primarily in terms of the accessibility of existing records and the solution in terms of the extension of user services. There is much that is valid in this representation but it gives insufficient attention to the important fact that accessibility is bound up with all of the production procedures and is inseparable in a number of fundamental respects from the issues related to the quality and scope of the existing records. It seems useful, therefore, to attempt a more precise formulation of the problem set to the solution of which the data center concept is addressed.

#### THE PROBLEM

The central problem of data use is one of associating numerical records. No number conveys any information by itself. It acquires meaning and significance only when compared with other numbers. The greatest deficiency of the existing Federal statistical system is its failure to provide access to data in a way that permits the association of the elements of data sets in order to identify and measure the interrelationship among interdependent activities. This deficiency has been partially overcome in a few vital areas where we need to trace and analyse the performance of the economy, by the establishment of special programs to bring together data sets in the form of national accounts, special index series, etc., but remains a debilitating constraint for most uses of data for analysis and planning. This is true for virtually all levels of use and for all purposes. It is a problem that plagues the research analysts inside and outside of the National Government who, for example, are engaged in building models of the economy in the interest of analyzing and projecting the major dimensions of economic growth and stability. It has been the principal obstacle to the administration's attempt to build a postattack revaluation and recuperation model. It is just as serious a problem for the uses that do not take the form of integrating data sets into a complex and formal model structure. The organization like EDA that wishes to establish a measurable test of eligibility for its program benefits faces the same problem. The business analyst who simply

wants to identify a variety of characteristics of the firms and households that form his principal markets often faces the same problem. The State Governor who wants to evaluate elements of his own program or the problems of his State by comparing them with the problems and programs of other States faces the same problem.

While the general problem is one of associating numerical records in use, it is understood better if we can visualize it in terms of its problem subsets. There are a number of distinct obstacles to file association that need to be identified before we can talk about solutions and program options. Consider the following:

#### *The archival problem*

This is the problem that initially interested the Ruggles Committee and the door through which they entered to their concern with some the broader issues of file management. The problem arises from the fact that the statistical agencies are oriented primarily to producing data publications and often leave their records improperly documented for further proceeding and analysis. Worse still, useful records are sometimes destroyed. These things occur because the existing system has no standards for identifying the files significant for preservation or for essential levels of file documentation. It provides no financial or organizational mechanisms for their maintenance. The decisions about the significance of archives is left to functionaries with little knowledge of their value in use and who must allocate funds for their documentation and preservation in competition with agency missions defined by previous policy in more restrictive terms and considered primary by agency personnel.

This, obviously, constitutes a major obstacle to the association for records for anyone who needs to work with data with any significant historical dimension.

#### *Problems of file maintenance*

Closely associated with the archival problem are some of the more fundamental problems of file maintenance. The utility of a file and its capacity for association with other records rests on more than the existence of a tape and a document that identifies its content sufficiently well for the data to be retrieved. Many additional problems stem from the low quality of file maintenance.

A couple of the more gross and obvious defects are associated with the fact that there are still important records that do not exist in machine readable form. Amongst the files that do exist on tape, some are in a mixed binary mode and some in a decimal mode making data association impossible without expensive and time-consuming mode translation. This is often true even between records of the same agency.

More subtle defects in file maintenance are uncovered when the need for data association requires bringing the detailed data in to accord with summary or published data. Often in the rush to meet production goals, agencies have pushed work through the processing stages of screening, reconciliation, estimation and summary in great haste without correcting prior files whenever errors or discrepancies are found. For example, corrections made "at the summary level are not carried back into the micro-detail. Indeed in some instances corrections have been made only in the published data, leaving both the summary and the detailed machine records uncorrected. Occasionally, summary data may no longer be in machine records and must be recreated by reprocessing the detail files or by keypunching and processing the new records. In some surveys, standards for screening data for creditability may have been coarser than appropriate for other uses.

The urgency to release results may also leave a disarray from the viewpoint of good file record and format management required for the files to be reused. There may be no uniform position in the records for like data and duplication of the same data can occur. For archival purposes a uniform record for the same data is essential and elimination of duplication economical. Also, the tape records may contain excess information. Certain codes and indicators used in the initial processing have no meaning in the archival context and the files must be purged of the excess information. Files may have interspersed alphabetic information useful only in the narrow survey context and which add complexity to programming efforts when used in other contexts. The absence of clear identifiers as part of the tape must be corrected to facilitate use of such tapes.

The requirements for simple access as well as the association of records are often stymied by the limitations of standards and procedures for routine file maintenance. Mr. Mendelssohn, who was loaned to OSS by BLS to conduct a

detailed review of the condition of the more important data files that might form an archival or file management center, concluded that the loss of data because of the failure to support good file management is distressing. (Appendix B.) The report of the study group formed at the National Bureau of Standards also emphasized the problem of file maintenance (appendix C) as did the Ruggles Committee report.

#### *The reference problem*

One of the serious obstacles to the fullest utilization of the information resources of the Federal Statistical System and to the effective association of its records in use, is the absence of any clearly defined reference function. The inadequate nature of this kind of service is directly traceable to the production orientation of the agencies stemming from their primary missions as data publishers. The reference function has generally been thought of as the responsibility of the documentary centers. To the extent that the agencies attempt to provide occasional referencing assistance, the task falls to an individual whose primary mission is defined in terms of the publication mission. The inadequate nature of this service is also traceable to the fragmented nature of the records of the Federal statistical system growing out of the divided responsibility for their generation and maintenance. The reference problem is made especially complex because of the decentralized character of the Federal statistical program. No agency has been in a position to perform a reference service in relation to the total file. The problem is becoming more acute as records are frequently demanded in disaggregate or special form not met by traditional documentary formats and are frequently used in ways that require extensive knowledge and understanding about the compatibility of records in several dimensions and the circumstances that condition their accessibility.

#### *The problem of administrative records*

Some of the most useful general purpose numerical records are generated as a by-product of administrative and regulatory procedures of Federal agencies. These agencies rarely interpret their missions to include the capacity to provide general purpose statistical services. Even when they might like to do so they experience serious difficulty because of the traditions, program priorities, budgetary procedures, and legislative authorities peculiar to their agency. Users who need to acquire from these agencies, and especially those who need to match these records with other files, find their task difficult. There is an important need to provide for the management and servicing of these records for general purpose statistical use.

#### *The disclosure problem*

The legal and administrative regulations on the disclosure of information supplied by individual respondents are becoming increasingly restrictive to the user. Only rarely is this because policy or research requires specific information about individual respondents per se. It is usually because of the need to associate sets of data in the interest of determining the interrelationship between two or more variables. The strain upon disclosure arises because matching several sets of data for consistency at levels of aggregation appropriate to the problem requires a retreat to elemental units in the process of constructing the necessary aggregates. It is not widely understood that the interest in micro-data and the existing pressures and constraints do not grow out of an interest in information about the specific respondent.

The fact that the strains upon the disclosure rules usually are of this form is fortunate because there are possible a number of servicing procedures based upon computer technology that can satisfy the needs of the user in most cases without violating disclosure regulations. Currently, however, the agencies of the Federal statistical system have only a very limited capability for performing the kinds of services that would lead to disclosure by-passes. The usage of the data is, thus, severely constrained, and valuable information is lost by aggregation at too early a stage in the analysis.

#### *Problems stemming from the procedures used to generate data*

It would be a mistake to conclude that the serious obstacles to effective use of the Federal statistical system under modern conditions is solely a product of its present inability to perform a series of user oriented services. Some of the most serious anomalies arise out of current practice in the production of data.

We have already seen how constructing legitimate disclosure by-passes forces one back to a manipulation of highly disaggregated components or even respondent units as building block elements. The collection and tabulation procedures of the agencies generate constraints on data use that lead in a similar direction. This arises out of some fundamental problems in dealing with the coding and classification of original source data.

As has been noted, a common form of information usage in analysis requires the matching of an attribute for two or more statistical sources or the association of two different attributes. This may require matching between different historical sets or between the files of different programs or agencies. This association of records is rendered difficult or impossible by at least three classes of record incompatibility that stem from current production practice. One difficulty in associating records from the different sources stems from the noncompatibility of classification of the data by several collection agencies and information systems. In the process of condensing and summarizing source records from initial respondents, information is tabulated on the basis of classification schemes that group items into classes. Often these classifications are inconsistent. (This is sometimes the result of the failure to develop general purpose standard classification codes applicable to all programs for these intermediate aggregates. It is often a result of the fact that standard codes are applied differently by different agencies so that there is no assurance that each agency (or program) will assign the same respondent to the same cell. Either of these cases often makes a comparison of the cells meaningless or difficult for purpose of analysis.)

A further difficulty grows out of the fact that the basis for classification applied by the collection agency in defining the cells may be inconsistent with the analytical or descriptive requirements of the user.

When either type of problem occurs, one solution is to return to the initial respondent unit or some other disaggregate building block and reconstruct consistent boxes of data. This yields the same class of by-pass procedure identified with the disclosure problem. At this point one may encounter another common problem in the form of the noncompatibility of the definitions of the respondent unit. This is a class of noncompatibility that not only renders questionable the comparison of seemingly similar cells for different systems (as in the other two classes of incompatibility) but may render difficult or even impossible the reconstruction of compatible cells.

The anomalies that grow out of these compatibility problems can be tackled in two ways. One often hears it proposed that general purpose standards for the classification of intermediate aggregates be considerably extended and aggressively applied to all agencies. This may not be the most fruitful line of approach. Existing standards may possibly be improved and made more general purpose by a more intensive analysis of user requirements and a concern with the issue is not unimportant. However, an attempt to force all uses into a common standards mold for intermediate aggregates has attendant disadvantages from the point of view of the user as well as the producer and the agency vested with the responsibility of formulating standards. General purpose classifications for intermediate aggregates always require some compromises in taxonomy that reduce the utility of the data for special purposes. Furthermore, a great deal of the data generated by the Government comes from programs that have a special purpose mission and restrictive legislative authorities and requirements that go with it. Forcing on these agencies a rigid application of general purpose codes for intermediate aggregates may be impossible and even undesirable because they conflict with special purpose missions. It does not appear to be a helpful possibility that all data sets can be arrayed in compatible boxes that will anticipate all uses. The attempt to deal with standards in this context will place impossible strains and burdens upon the machinery for making and enforcing policy with respect to standards.

A more fundamental way to handle this problem may be a progressive move in the direction of compatible building blocks that can be reassembled to provide compatible and relevant aggregate sets for special uses and can be used as a bypass for disclosure problems and other procedural obstacles. This suggests that the problem of standards of greatest importance in the emergent situation is the need for uniform identification, definition and coding of the respondent unit as a basic building-block unit. The absence of a uniform system of coding and classification for geographical areas is also a serious deficiency and is an important part of this same problem. This also suggests the importance of procedures for assuring that every agency puts each respondent in the same cell

and that an important criterion for evaluating existing standards for intermediate aggregates is the extent to which they can serve as useful intermediate building blocks that obviate the necessity for returning to the respondent unit for many programs.

Unless something of significance is done to modify current practice in these production procedures, the matching of data from diverse sources will remain generally impracticable and often impossible.

#### *Associative records that do not exist*

Many of the most important analytical and policy issues of today require the association of existing records with records that do not exist. There are serious gaps in the public record of social activity.

There is a particularly important class of records that is missing. It can be identified by examining the problems of some of the most important Federal programs. Missing are the records that enable policymakers and planners to understand adequately how people, households, regions, activities, enterprises, and administrative units are functionally related and how they change over time.

The importance of such a capability is readily apparent. There is a large array of new and old welfare programs involved in trying to ameliorate various forms of social pathology and transform people (e.g., poverty, education, health), and regions (e.g., EDA, rural redevelopment), and the activities that engage them (e.g., Small Business Administration and large elements of the agricultural program). There is a large array of new and old programs engaged in planning for and providing public facilities (e.g., highways, mass transportation, water resources, urban development and housing, etc.). In each of these programs considerable effort, planning, and resources are expended for program development, in establishing the formulary for program management, and in evaluating program results. Indeed, Federal legislation in these areas impose planning requirements as a condition of grants-in-aid and other forms of assistance upon State and local governments and other State, local, and regional activities in at least a dozen large programs.

To date, the problems have been formidable and the results unimpressive for one principal reason. The information base that exists and can be economically accessed tells us a great deal about the characteristics of people, households, activities, enterprises and their institutions at any one point, but tell us very little about how they are linked into functional networks or how they transform over time. These latter are the most relevant information resources for policymaking and program evaluation in these areas. What form of job training, what form of regional assistance, what kind of road networks, what modes of mass transportation, what kinds of cities are questions that need to be answered on the basis of some knowledge of functional linkages and evaluated in terms of measurement of change. The responsible planners and administrators of these programs are feeling a keen sense of frustration because of the paucity and irrelevance of much of the information available to them. Some of the records they need to associate to resolve these issues do not exist.

The information gap related to these requirements reveals two elements of significance for the present evaluation:

First, a large part of the apparent gap in the kind of information needed is a direct function of the same system anomalies outlined above. In many cases, the problem does not rest upon the fact that the relevant attributes of people, activities, or institutions are not included in existing records. It rests with the fact that these attributes cannot be associated in functional configurations or traced through a historical sequence. We see the evidence of this in the widespread current interest in what is characterized as "longitudinal" data—(the ability to trace attributes of the same respondent through time in order to identify transformations—i.e., from and to movements in relations to places, activities, occupations, institutional affiliation, welfare categories, etc.). Therefore, many of these requirements could be met if the problem of file compatibility could be resolved through an extension of the servicing capability and some modification of the production practice of the Federal statistical system. This underscores the observation made in the introduction that the problems of file accessibility cannot be successfully separated from all of the issues related to the quality and scope of the files. In a fundamental way, file accessibility is the issue of file compatibility which is inseparable from the production practices that determine the organization and quality of the file. It is important to note that steps that can be taken to improve file compatibility



and accessibility will substantially increase the effective scope and utility of present files without a change in the size of the files or the attributes of the respondents they contain though additional resources might be required.

Second, although a significant part of the gap in information is a function of file incompatibility, there are also gaps which result from missing attributes.

For example, some attributes of the population may appear only infrequently with the decennial census and be needed for intercensal periods for vital program planning and evaluation. Other attributes may not appear in any records under existing information programs. Some notable gaps appear in the fields of transportation and construction and in connection with some of the important welfare attributes of people.

The following observation is pertinent here: The system, as it currently operates, provides no authority or mechanism for the review of the statistical program and the allocation of its resources in the light of the most important changes in information requirements. The decisions about programs that determine the scope and quality and accessibility of the records are primarily made upon the basis of technical problems, cost considerations, respondent pressures, etc., that impact directly upon the production process of the individual program and agency. There is no systematic way for the requirements side of the problem to enter the decision process.

#### *Potentialities for system development are being missed*

The new technology is making feasible a number of possibilities for greatly improving the utility of existing records. There are new possibilities for more efficient management of large-scale numerical files in terms of storage and retrieval; new possibilities for rearranging files in more useful form; new possibilities for retrieving in the form of maps, graphs, charts, and other media in addition to the traditional tabular forms; new possibilities for building in disclosure controls and disclosure bypasses; new possibilities for matching records to assure compatibility, etc.

One aspect of the service potential inherent in the new technology deserves some elaboration. The association of records in analysis usually carries with it a computational burden. This may take the form, for example, of computing the ratio of two data sets or making seasonal adjustments or computing coefficients in the analysis of variance. All of these derived numbers form a latent set implicit in the original source data. The computational capacity of modern computers is such that computations of this kind can often be made as fast or faster than the tape can be passed through the machine. Once a system has been developed for providing such a service, the marginal cost of generating these numbers when the tapes are being passed for retrieval is close to zero. Adding this kind of system capability can have the effect of increasing the effective size of the files of the Federal statistical system tenfold with latent numbers involving some computation.

These potentialities require the expenditure of time and effort on system design and software accumulation that few agencies can justify in terms of their current program levels or even appreciate in terms of their existing individual program missions.

#### *The problem of file fragmentation*

This is not a problem that is conceptually distinct from the others. Instead, it intersects the entire problem set being discussed and forms a part of the explanation for some of these anomalies. Currently, files are being generated and managed by more than 20 different agencies. It is precisely this division of responsibility and fragmentation of resources that inhibits system development and generates many of the problems of file compatibility. But apart from the way this problem invades all of the others, it imposes additional constraints because of the procedural, bureaucratic, and sheer time and space restrictions upon file usage.

#### THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability, accessibility, and compatibility are very high. This is apparent even if we restrict our view to the significant Federal programs mentioned above. There is manifest in these programs an impressive and unexploited joint demand for information. This jointness has two important aspects. Even where the attributes of the numerical files of importance to these agencies are disjoint, they require the same servicing capabilities in the statistical system in order to

perform the essential tasks of record association. The agencies require the same kind of system capabilities. Beyond this, a number of the programs have a joint interest in the same sorts of file extensions. The new welfare agencies, for example, have a strong joint interest in longitudinal data about the welfare attributes of people that are not currently available. Even where these agencies might have discrepancies in the attributes of specific interest, there is a good possibility that the same collection vehicle could be used in servicing their needs.

It is interesting to note that many of these agencies have had substantial resources given to them by Congress explicitly for the purpose of generating or accessing the information essential to the conduct of program. This constitutes a formal recognition by the administrative-legislative process that the established statistical programs are not meeting these needs. So far, the remedial choice has been to fund programs to meet special requirements rather than system modification. This kind of bypass, however, has proved largely ineffective for several reasons.

(1) These agencies have no effective way to apply these resources to system reform that would improve record compatibility.

(2) The funds are dissipated because, though significant in total amount and perhaps even adequate to support major improvements, they are fragmented by their attachment to specific and narrowly conceived missions with not appreciation of the overlap or jointness of interest.

Thus, while the new welfare agencies could probably finance a collection vehicle adequate to their joint requirements, no one agency can really accomplish this satisfactorily alone and there exists no coordinating authority that can identify and exploit their joint interest.

Further fragmentation occurs even within agency programs. The Corps of Engineers, for example, has for years spent enormous sums of money on information to serve water resource planning requirements. Much of this expenditure has been duplicative and wasteful because the money for this purpose has always been funded on a river basin project basis so that it was virtually impossible to take advantage of the scale economies for building the servicing capability for the entire set of projects. As a consequence, each river system has tended to be planned in functional isolation without the opportunity to define the linkage between projects or to trace economic and social costs and benefits in an appropriately general context.

The stakes in program improvement in the Federal statistical system are high because the amounts of money being ineffectively spent on statistics in these programs is very large. They are also high because the improvement in the utility of the information base could have an unmeasurable but substantial effect upon the quality of public administration. The ability to ask relevant questions and get prompt relevant answers in planning, administering, and evaluating programs is of considerable importance.

All of this is only by way of recognizing the Federal interests involved. The stakes of State and local public officials, and the business and research community are equally large. These are the decision units which, by the nature of their responsibility, require disaggregate data sets that are especially affected by the problems of file compatibility. They have a common interest in extending the capabilities of the Federal statistical system.

This common interest has an especially important new dimension. The major opportunities that exist for extending the scope of the file available for analysis with some reasonable economy of effort in the near future rest in an exploitation of the records that are (or can be) generated by the State and local public agencies. However, their utility, and the utility of the file of the Federal system, will be immeasurably enhanced if these records can be brought into reasonably compatible association.

We are witnessing a burgeoning interest on the part of the State and local groups in developing the mechanisms for setting standards for these files and for maintaining and servicing them over time. This source of information is bound to emerge in importance and size. It is particularly important, therefore, that improvements in the Federal program lay the groundwork that will permit effective integration of the Federal file with other sources as they emerge. Furthermore, these emergent efforts are going to require guidance and leadership in setting standards and designing systems in a way compatible with total requirements. Much of this leadership must be supplied by example and by cooperative effort by a Federal system that is moving in response to modern requirements and opportunities.



## THE SOLUTIONS ARE COMPLEX

It seems clear from the foregoing problem characterization that the solutions to this set of problems will have to be multidimensional.

When it made its proposal for a national data center as a solution and developed its justification, the Ruggles committee revealed an understanding of many of these dimensions. Its proposal was a constructive one and intended to be interpreted with some flexibility. However, the representation of the solution in this way has had some unfortunate consequences not anticipated by the committee.

In designating the center as a national data center and placing considerable emphasis upon the collection of tapes—growing out of its concern with the archival problem—the proposal became quickly translated in the minds of many as another data bank proposal.

The data bank idea is enjoying a considerable fad at the present. Many people have grasped this as the solution to their information problems. They have been encouraged by the substantial success that some fairly restricted and specialized information systems have had. A number of businesses, for example, have enjoyed some success in pulling their management records into a compatible and useful information system. The impression is widespread that bringing machine records together into some kind of central file will be instrumental in resolving the data problems of the broad class of users who attempt to use the files of the Federal Government.

This notion is supported by a general misunderstanding about the character of the files and their use. Those supporting this view are impressed with the fact that many different users have intersecting requirements for the same sets of data produced by the Federal statistical system. These records are, accordingly, viewed as general-purpose files. The convenience and economies of scale of bringing these records together into a common repository seem obvious. The obstacles to effective use under the present system are interpreted as technical and bureaucratic limitation amenable to this kind of technical solution.

What is not often adequately appreciated is the fact that general-purpose data are always used to fill special-purpose needs. This means that, while there are many intersecting interests in the same files, the impact on the file of each use may be quite different in terms of the organization, the levels of disaggregation required, and, most importantly, in the way the file needs to be associated with other records. It is this need for record association that is paramount and the source of most of the difficulty as was represented above. What makes a record a general-purpose record is not the fact that many users have an interest in its dimensions. It rests upon the file being constructed on the basis of standards, maintained in effective condition, and serviced by institutional arrangements and a technical system capability that will allow it to be reprocessed and combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users.

Thus, the key to solving these problems does not reside in the assembly of the records in a center but in the capacity to provide certain forms of file management and utilization services to the user. The effective provision of these services may require the assembly of some of these records into an integrated file, but this is defined by technical system requirements and is not the central issue it is made to be by many representations. It is important to characterize such a program as a data service center. The proposal is too important and fundamental to be burdened with its association with the naive data bank concept.

The Ruggles committee explicitly formulated at least a part of this rationale in their report and, hence, were putting forward a constructive proposal worth serious consideration. However, the committee never made explicit the way in which the problems of file compatibility rest upon the collecting and tabulating procedures of the agencies. It needs to be emphasized that these are important dimensions of the problem. Extending the mission of the Federal statistical system to provide user servicing capabilities based upon the new technology can do a great deal to extend the utility of existing records. However, the logic of a flexible service capability rests upon the ability to manipulate statistical building blocks. The development of these building blocks is a production task not contemplated in the suggestions for a data center. Some modification of current production practice will be essential for success. Indeed, if this problem is not tackled on a broad front, the generation of the servicing capability will fail to provide the kind of service intended and aggravate the sources of friction and dissatisfaction vis-a-vis the producing agencies.

## RECOMMENDATIONS

The concept of a national data center is an appropriate vehicle for program reform if the concept is broadened to emphasize the role of the servicing capability and if it can be given an important role in assisting the Bureau of the Budget establish standards and monitor compliance. Accordingly, the basic recommendation is for the establishment of a national data service center whose primary mission would be to provide service to users of Federal statistical data both inside and outside of the Government.

This Service Center would have to be designed from the outset to incorporate certain basic functions:

(1) Direct the file storage and management for significant archival records in machine readable form for all participating agencies.

(2) Provide a central referral and reference source for the users of Federal statistics. This would include the development and maintenance of a formal reference index and the development of statistical reference specialists.

(3) Provide explicit facilitating services for the users of Federal data. This capability would consist of the following kinds of services:

(a) File rearrangement and cross tabulation to meet special needs and provide an extended range of output options in the form of maps, graphs, charts, and other media in addition to traditional tabular forms;

(b) Tape translation and other forms of file modification to bypass some of the inconsistencies and deficiencies in file management;

(c) Record matching where file compatibility exists or can be developed by file rearrangement;

(d) Disclosure bypassing where requirements violate legislative or administrative regulations;

(e) Perform standard statistical routines that form an essential part of the strategy of record matching and disclosure bypassing and which join routine computation with retrieval in a manner that makes a whole set of computationally derived numbers a latent part of the file of the Federal statistical system.

The National Data Service Center would be prepared to perform these services for:

(a) Archival records under direct management control of the Service Center;

(b) The current and accumulated records of administrative and regulatory agencies; and

(c) As a system resource or facility available to be used in connection with the current records of any agency where the need cannot be adequately met by the agency.

(4) Develop the computer hardware and software systems essential to the file management and servicing functions.

(5) Provide the staff support to work in conjunction with the Bureau of the Budget to develop and establish standards essential to the system capability.

There are a number of areas in which new or revised standards will be essential:

(a) Standards that define the records to be preserved in archival form;

(b) Standards for documentation and file maintenance, and

(c) Standards for the classification and coding of statistical data with special attention to respondent units and other forms of statistical building blocks.

(The sooner some of the standards related to the establishment and maintenance of archival records can be established the better. The review of the program [in app. B and in the next section] indicates that the most useful archives and the most economical are those that are developed under proper control and coordination from the present forward. The necessary procedures can then be built into the routine processing of data. This suggests some urgency for making as much headway with these issues as possible. The OSS should begin right away to work on establishing these standards without waiting on any formal actions on proposals for a data service center.)

(6) A research-analytic capability will be essential to the success of these functions. This does not mean developing the capability for conducting research and analysis directed toward issues of policy and management. Such analytic functions should be centered in the Executive Office and the operating departments. Policy research and analysis should be kept separate from the supporting function of supplying and servicing information.

A research capacity directed to an analytical evaluation of user requirements for the purpose of designing and developing the system components essential to perform these services is the essential capability. The construction of the reference file, the definition of standards in every category and the design of software routines and other system techniques that perform the facilitating services would all be controlled by what could be learned through research and analysis about the systemic elements of user requirements.

Some indication of the direction this analysis will have to take can be gathered from the National Bureau of Standards report in appendix C. A modest effort made to think through the kind of knowledge about user requirements that will be essential to system design and development is described there. In sum, it will be important to identify major classes of users; to learn the extent to which their requirements intersect the same sets of data; to learn the ways in which they require record matching from similar or different sources and the acceptable levels of aggregation. Only a systematic understanding of the joint and disjoint characteristics of the major requirements can serve to design an effective reference index, design relevant standards, and guide system design.

#### RESOURCE REQUIREMENTS

Many of the elements of this kind of program appeared in the Ruggles committee report and in the preliminary review. The task that has occupied recent months has been the attempt to document the needs more fully and develop some notion of preliminary specifications and costs.

The problem was broken into three parts for study and discussion and assistance sought with each. First, the essential ingredients for a reference and referral service were considered. A committee of knowledgeable people was assembled on an informal basis to discuss these issues (identified with the report in app. A). Second, a more intensive study was undertaken based upon the survey of machine-readable records conducted by the Office of Statistical Standards and contained in the appendix of the Ruggles committee report. In this way, an attempt was made to specify more clearly some of the costs of the archival function. The Bureau of Labor Statistics made part of the time of Mr. Mendelsohn available to carry this out (report in app. B). Third, an attempt was made to specify more clearly the essential elements of the system that would provide the facilitating services and what it would take to provide such services. For this purpose the National Bureau of Standards was used as a vehicle to assemble several people with a considerable range of knowledge of both the uses of Federal data and the production processes that generate them (report in app. C). In addition to these organized efforts I have discussed substantive issues with a number of knowledgeable people in the Federal agencies (both statistical agencies and program agencies) and in the universities (including an interview with the professionals involved in Project MAC at MIT).

This effort has yielded a better understanding of the nature of the problem and the system requirements. However, it has been somewhat less successful in specifying in detail the components of the system and the resource requirements. Let me review first the results and then evaluate the shortfall.

#### *The reference function*

In reviewing the requirements for the reference function the committee made a rough judgment that it might take as much as 5 years and an average of \$2 million a year to provide a meaningful reference and referral service for the Federal statistical system. This appraisal is limited in two ways, however. It is not the product of the kind of staff work in program planning that would be essential to a refined estimate and, therefore, represents only an informed speculation. More important, this estimate was generated with a view of the reference function as a discrete service unit or capability. It is recognized that a reference service would be more effective as an integrated part of a total service center program because the reference problem forms only a part of the larger problem set. If the provision for reference services is combined with other user services, the professional staff (particularly in its analytical and system development capacity), the computer facilities and other components of the service system could perform many joint functions. It is believed that because of the joint product character of these services, the incremental costs of providing a reference capability as part of a larger service system would be somewhat less.

#### *The archival function*

A review of the tape file inventory was undertaken which attempted to identify the important archival records and determine the costs of bringing these files to an acceptable level of file maintenance and documentation to be incorporated into an archive. The question of which records constitute significant archives rests, of course, upon an interpretation of requirements and development of standards not yet undertaken. In the interest of getting some feel for the dimensions of the problems this question was begged by arbitrarily preparing a list of the records considered to be vital general purposes series on a judgment basis to represent a sample archive.

On the basis of the data included in this sample archive it was estimated that a more complete archive would represent about 20,000 reels of magnetic tape and require an estimated \$3 to \$3.5 million and 3-5 years to develop. Of this amount about \$800,000 would be needed to bring data not now machine accessible into usable form, about \$500,000 would be needed to transfer punched card data to magnetic tape. In addition, between \$500,000 and \$1 million would be essential for blank reels and tape copying.

One of the interesting aspects of this report is the fact that almost half of this file (9,000 reels) could be brought into a data center for about \$260,000 within a year. This indicates that the files vary widely in the quality of their maintenance and documentation. The incremental costs of the second half is about \$300 a reel as compared with \$37 a reel for the first 9,000 reels.

Obviously, the costs of bringing existing files into archives are substantial and some review and justification will be needed. This can only be done within the context of a more comprehensive review of the user requirements that must guide planning in this area. A related sidelight of the report is that the files that are best maintained and can most easily be brought into an archive are not necessarily those that are most important in a usage sense. This is largely a function of the size of the files and the frequency with which it is produced. This suggests that an archive based upon considerations of cost and convenience in assembling existing records does not assure the most useful file.

At the same time, a large part of these records can be preserved at such a small cost that there seems little question that the investment in this resource is essential and justifiable. There is another important observation. About half of the total costs estimated are for system development and will have to be incurred even if the archival objective is addressed to current and future records only. These must be thought of as a capital cost of archival development as distinct from the costs of "dusting off" existing records.

It is true that these records will still contain all of the elements of file incompatibility that are the product of the production methods and standards that governed their generation. However, they appear sufficiently vital to current and future analysis that a total program should make a serious provision for trying to salvage some of the loss in data resources that has taken place in the absence of a policy and procedure for file preservation. In addition, every step needs to be taken to place future accumulations on a sound and economical basis.

The costs of bringing these records into an archival file do not represent all archival costs, of course. There are storage costs also (less than \$10,000 annually for 20,000 reels in prime air conditioned space), but these are inconsequential when compared to the need for facilitating services resting upon hardware and software systems to allow their effective use. These latter costs, however, cannot be fairly estimated at this point because, again, these services can be provided jointly by a facility which has a broader user service mission. As in the case of a reference service, an archive that is set up as a discrete service function will cost more than one incorporated in the total service complex.

#### *A systems capability for facilitating services*

It was through the agency of the Bureau of Standards that we attempted to assemble the intellectual resources to establish the scope of the program and the costs that would be required in establishing a system capability for providing the kind of facilitating services outlined in the recommendation. It is obvious from the foregoing discussion that this is the key to the program concept and to the evaluation of costs.

This turned out to be a difficult assignment. In our early attempts we found it exceedingly difficult to estimate program dimensions and costs without a clearer specification of the requirements the system will be designed to serve. Accordingly, a 3-day study session at Camp Ritchie was planned to see if we could break the back of this problem as the report in the appendix reveals.

This session made some progress in visualizing the requirements as well as assisted in clarifying some elements of the problem to be solved. In particular, we began to formulate some notion of how an analysis of user requirements might be structured and carried out. We could not, in the time spent, get to a more precise specification of the system elements. I feel that this work has laid the groundwork so that a series of additional work sessions of the same type might lead to a formulation of more specific program options. This task was found to be large enough that it did not seem advisable to undertake this kind of intensive staff work without a more specific decision on the part of the Director of the BOB concerning the kind of effort that is going to be devoted to this problem set and under what auspices.

At the same time some informed speculative judgments about the order of magnitude of costs developed out of the discussion. A total program of the type outlined under the recommendations would probably start out with an expenditure of \$1 or \$2 million annually in the first years and grow to the neighborhood of at least \$10 million a year. In the early years the size of the program will be controlled more by the practical limitations of assembling and training the kind of staff and acquiring the kind of equipment necessary. The fund could not be spent at a rate commensurate with the need and the objectives. A more detailed specification of program objectives, their phasing and the allocation of costs will have to rest upon additional staff work and should be preceded by some tentative policy decisions that will guide the work.

#### *Staff requirements*

Just as it is not possible to detail costs, it is not possible to detail personnel requirements. However, all of the reports or the discussions leading to them emphasized one point. The kind of statistical reference specialist, user service specialist and statistical systems analyst that is required to make this kind of program work either does not exist or is in extremely short supply. This implies (1) that program development will be constrained at the outset by intellectual resources and not financial resources, and (2) a successful program in this field will have to make explicit provision for professional development and training both in its program and in its budget.

#### *Technical requirements*

This is also the place to point out that there has been nothing in the entire review to suggest that an effort of the kind outlined in this report would be technically constrained. There has been some indication that existing computer hardware has been designed with greatest concern for computational capacity and is not as economical or as flexible as possible for the management and servicing of very large scale numerical files. However, the existing state of the art contains the essential elements of a more appropriate hardware system.

Similarly, the software routines for file management and servicing will need to be developed, but there is nothing to indicate that these problems of system development are not tractable. What is indicated is that considerable work must be expended over time to create these capabilities. There is every assurance, however, that the state of the art is adequate to support fully this kind of effort.

#### ISSUES OF ORGANIZATION

My views on the organization issue have been strengthened by the months of study since the preliminary memorandum. First, I cannot visualize a meaningful program addressed to the interrelated set of problems discussed above without a considerable degree of centralization of function. Some form of inter-agency service center will be essential. Second, if such a center is developed with existing agency structure essentially unmodified, it cannot perform its mission without agency cooperation and without explicit accompanying modification of agency missions.

This suggests that an effort of this scope could not be implemented without seeking new legislative authority. Legislation will have to be worked out and sought that will permit the service center to receive file custody, that will relieve the agencies of their disclosure restrictions as they pertain to the release of data to the center, and, at the same time, transfer the agency's disclosure obligations to the center. No workable independent center could be developed without meeting this issue head on at the outset. Further, the kind of program coordination and control of standards that will be essential may require legislation giving some interagency program authority to the new center. In addition, a single budgetary instrument for implementing the new program would be essential.

This kind of formulation inevitably leads to speculation about the organizational forms that might serve this end. I would like to react to some of the speculation proposals known to me:

(1) A new independent agency; this is certainly the cleanest solution. It could be accomplished with a minimum of "ad-hoc-ery" and would provide maximum flexibility for planning and innovation. It would be unencumbered by many existing agency jealousies and provide the freest opportunity for developing new leadership. At the same time, it might foster a coalition of agency opposition. Given the attitudes both in the Bureau of the Budget and on the Hill about new independent agencies, it might be difficult to manage. Barring this, some existing agency would have to form the vehicle for the program. Several have been mentioned in this context.

(2) GSA-Archives. The General Services Administration and, specifically, Archives, has been pointed to as an agency with already existing interagency authority and concerned with a part of this function. I am extremely dubious about the viability of such a solution. In the first place, this is not primarily an archival problem. It is primarily a complex problem of file management and coordination and rests upon a base of production practice that must be involved in the solution. It must develop a mixed professional staff of reference specialists, statistical specialists, subject matter specialists, system design specialists and programmers and technical services staffs. These resources do not exist even in embryo in these agencies. They would be handicapped by their image in building up the quality intellectual resources the program would need to succeed. Furthermore, these functions would not represent a primary mission from the point of view of the management of these agencies. An even more compelling objection is the fact that it would take the first tentative steps toward some integration of the user services of the Nation's statistical system down an organizational path that might make more difficult the achievement of desirable emergent forms.

(3) One of the existing statistical agencies: I do not believe that any of the existing agencies offer a desirable home for this function. It is true that they have already accumulated some of the expertise and equipment and management services and experiences that a new venture of this type requires. However, their mission concepts are conservative and inhibited in this area. The leadership is lacking. Perhaps more important, old interagency jealousies, etc., would make it more difficult to develop an atmosphere of cooperation.

(4) Compromise between the existing agency and independent agency solution: Of the existing agencies Census is certainly the most logical candidate by virtue of both its mission and the caliber of its professional staff. Many of the disadvantages of assigning this function to a new agency would be offset if Census were made an independent agency itself. If this were done and the user service functions set up parallel and with equal organizational status to the Census function we might have something of the best of both worlds. Something of this type might also have the advantage of being a constructive first step toward some degree of reorganization of the Federal statistical system.

(5) The National Bureau of Standards: It has been pointed out that the Bureau of Standards performs an interagency mission and has an unusual combination of existing legislative authorities to receive funds from and distribute funds to agencies, to set up special institutes, to use visiting scholars, etc. It already has an interagency service tradition and has been explicitly given the responsibility for assisting Federal agencies in planning computer systems. They also have in existence an emerging computer utility that might serve some of the needs.

(6) National Resources Evaluation Center: This agency has been suggested because it is an independent agency in the Executive Office of the President with existing interagency authority and responsibilities that extend beyond the mission of the Office of Emergency Planning where it is housed. It already contains a very large file of integrated Federal data from the various statistical agencies and has built up a staff with more experience in integrating interagency records than any other agency. It has a large computer installation organized for large-scale file management. It has an already existing interagency committee with the major statistical agencies represented and each of the agencies have one or several professional employees assigned full time to the activity. It is a conceivable vehicle if its authorities and functions were extended and removed from OEP.

If an effort to implement such a program goes forward, a great deal of thought and discussion will have to go into reviewing the kinds of options represented

here. I would like to offer the following related observations: First, there is a major threshold that must be negotiated if we deal meaningfully with the problem and program set at issue here. The kinds of services recommended cannot be subdivided without imposing upon a more limited function, serious functional handicaps, unnecessary expense and, possibly, seriously threatening its chances of success. I am concerned that partial measures may, in the end, do more harm than good. Since such an effort cannot spring into being "full-blown," it will have to be time-phased over a considerable period. However, the program should be considered as a whole.

Second, from an organizational point of view it seems inescapable that whatever initial action is taken, the end result will be a substantial reorganization of the Federal statistical system. It is very important that the organizational vehicle used at the outset does not predetermine the future evolution of the system in a way that limits its ability to implement essential subsequent phases.

Third, it seems to me that in discussing this problem set and proposed program, we are really engaged in a discussion of what kind of Federal statistical system we want to develop in the next generation. I am much less sanguine now about whether it is possible, or even desirable, to keep these issues in the background. I am sufficiently concerned about the abortive potential of solutions that fall short of a critical threshold and organizational arrangements that might inhibit essential lines of development that I feel it important to consider the issues of organization in a straightforward way. I would go further and suggest that every one of the six interim solutions outlined above has serious limitations, although some have a more open-ended character than others. My own preference would be to handle the organizational problem at the outset as a reorganization of the Federal statistical system. There are a number of indications that this might be a favorable time to do so.

Fourth, this predilection is reinforced when I reflect upon the great importance production practice plays in the whole configuration of problems and solutions. I am also inclined to believe that a fundamental improvement in the integration of production practice can offer a constructive solution to the paperwork problem of respondents without jeopardizing important components of a general purpose information system. Another factor reinforcing this inclination is the conviction that some form of integrated leadership can go far in dealing in a creative way with the joint interest of existing Federal programs and agencies whose current large expenditures for data now constitutes a large resource waste.

#### PRIORITIES

The comprehensive scope of the issues presented here plus the fact preliminary staff work cannot specify and cost explicit program options suggests certain priorities. First, a continuous focus of leadership needs to be generated. The proposal has already been put forward that an interagency committee be established to provide this focus.

Second, wherever the leadership function is vested, it seems to me that the highest priority is to provide this focus with the staff support essential to identify requirements and specify the elements of the system that must be provided for. The earliest requirement is to engage the research-analytic capability identified under item 6 of the recommendations. The development of specific program options, the definition of their specifications, costing these elements, and identifying the essential order of a time-phased program will require early intensive staff support of a very special kind.

Third, a beginning can be made under existing authority upon some of these problems before new programs and organizations can be developed and funded. At least two kinds of effort could be begun right away.

(a) The standards that shape the content of archival records and determine the essential forms of file maintenance and documentation need to be worked out and made a part of ongoing programs. A beginning can also be made in formulating the kinds of standards that will produce statistical building blocks essential to file compatibility.

(b) The 9,000 tape file record identified in the Mendelssohn report constitutes a nuclear archive that can be generated quickly at a very modest cost. Funds should be made available to the agencies to begin the creation of this basic archival record.

The Bureau of the Budget should seek funds to carry out these preliminary measures under its own authority. The staff work and the extended effort applied to statistical standards should be centered in the Office of Statistical

Standards. The funds to develop the basic archive could be transferred to other agencies as a part of a controlled plan.

Several hundred thousand dollars might profitably be requested in the fiscal 1967 budget for this purpose.

#### APPENDIX A

##### APPENDIX MEMORANDUM

Subject: Report of informal committee on the reference problem.

#### THE COMMITTEE

An informal ad hoc committee was assembled by Edgar Dunn, acting as chairman, to discuss the problem of developing an adequate reference service for the Federal Statistical System. The committee was composed of Joe Daly and Edwin Goldfield of the Bureau of the Census, Rudolph Mendelssohn and Robert Steffes of the Bureau of Labor Statistics, Ezra Glaser of the Patent Office and Edgar Dunn, consultant to the Office of Statistical Standards. These individuals participated with the knowledge and consent of their agencies but served as individuals and informed professionals. No attempt was made to get agency clearance or establish agency points of view. It was merely an informal attempt to formulate some judgments about the nature of the problem and its solutions. The committee met on several occasions on an irregular schedule during the summer of 1965.

#### THE PROBLEM

It was agreed that one of the serious obstacles to the fullest utilization of the information resources of the Federal Statistical System is the absence of any clearly defined reference function. The inadequate nature of these services is traceable to the production orientation of the agencies stemming from their primary missions as data publishers. The reference function has generally been thought of as the responsibility of the documentary centers. To the extent that the agencies attempt to provide occasional reference assistance, the task falls to an individual whose mission is defined as a production responsibility. The inadequate nature of this kind of service is also traceable to the fragmented nature of the numerical records of the Federal Statistical System growing out of the divided agency responsibility for their generation and maintenance. Because of the decentralized nature of the Federal statistical program the referencing problem is made especially complex and no agency has been in a position to perform a generalized service with reference to the total file.

The problem is becoming increasingly important in recent years as important uses of the numerical files are more frequently taking the forms of records in machine readable form rather than the traditional documentary form. The problem is also fed by changes in information usage that are leading to more complicated information requirements. Records are needed more often in disaggregate or special form not met by traditional documentary formats and they are often used in combination in ways that require extensive intelligence about the compatibility of records in several dimensions.

#### DESIRED REFERENCE CAPABILITY

Ideally the Federal statistical system should be able to develop a reference system that has the capacity to deal with inquiries in an efficient and creative way that would facilitate access to the records and extend their utility. The clientele is conceived to cover a wide range of sophistication and types of need. However, the requirements for a reference capability stem primarily from a large and growing core of intermediate information processors that service the research aims of academia and the decision and administrative requirements of business and government at all levels.

Such a service should be able to:

1. Help the client refine his inquiry and frame it in a way acceptable to the system and, in the process, give some preliminary information about the scope and nature of materials implied in his request as an aid to further defining, sharpening, and limiting the inquiry. (Experience of the Science Information Exchange and the National Science Referral Center have pointed up the great importance of this function even in dealing with highly trained professionals.)
2. Provide, by drawing upon a reference index and other reference tools, a fairly complete documentation of formal intelligence concerning—



- (a) the number and size of relevant file sets,
- (b) their taxonomic descriptors,
- (c) their mode of preparation (census, survey, etc., questionnaire forms, etc.),
- (d) their quality characteristics including (1) quality of the data (sampling and response errors, etc.), and (2) quality of the files (state of documentation and file maintenance),
- (e) the extent to which the taxonomic and qualitative characteristics of the data will support merging and collating series for various purposes,
- (f) where the data are located and how they may be accessed including such information as the form of the file (published, machine readable, machine language and format, etc.) and access costs in both time and money;

3. To perform a switching service so that the user can access the needed records efficiently. (The uninitiated user needing to access several data sets in different divisions of several agencies can be faced with a complex switching problem difficult to handle efficiently unaided.);

4. To provide a "semiautomatic Joe finder" to facilitate access to informal and specialized intelligence concerning the records and their characteristics. This would need to be a person-oriented service that would revolve around the role of the professional reference specialist who would deal with the client by person, by mail, and by phone.

#### IMPORTANT CONSIDERATIONS IN IMPLEMENTATION

A general consensus emerged from the committee discussion about the principal issues or problems to be resolved in the process of implementing such a goal. There was also agreement about the general form of the solutions. The most important issues revolve around three points:

##### *The reference file*

In order to perform his role effectively the reference specialist would have to have access to a set of formal reference aids that would constitute the elements of an emerging reference system. The principal aid is visualized as a formal reference index that would probably be machine oriented. This index would attempt to bring into a reference file the kind of reference intelligence implied in the previous sections (2 a through f) that could be gathered and formulated in a formal system. Such a file would be designed to facilitate an iterative search procedure and to generate documentation to service each inquiry. After some initial period of development this file might have the capacity to generate, periodically, one or more condensed summary index documents that could serve as visual reference aids not only in the reference center itself but in documentary and service centers throughout the United States.

The construction of such an index would be a professional task of considerable magnitude and complexity. It would take time and resources to develop and maintain on a current basis. Indeed, the development of such a file would represent an ongoing task that, by its very nature, would never be completed.

The order in which the components of this reference intelligence are selected for development and the form of their organization into a file should not be random but governed by systematic priorities. Furthermore, it should be only marginally controlled by the ease with which such reference material can be organized out of existing materials. The development of the file should be controlled by a research-analytic effort on the part of the staff that would provide guides to the emergent usage. Accumulating knowledge about request incidence will be only partially helpful. Maximum effectiveness of the file will rest upon an explicit effort to identify the principal classes of users (in terms of their analytical requirements and problem orientations), the way in which their requirements are common or disjoint, the way in which they generally intersect different statistical records, and, therefore, the nature of the reference intelligence necessary to serve each class of user. Some idea of the systematic character of the requirements is essential in order to do an effective job of designing a responsive reference system.

##### *The reference specialist*

The reference file is only a tool. The key to successful data referencing is the reference specialist.

In the context we are discussing here he has three recognizable functions. First, the professional reference specialist is the essential human link in deal-

ing with the reference client. He must often assist the client in refining his inquiry to a form that will facilitate response. He will provide the interface with the formal reference tools such as the reference index and the "semi-automatic Joe finder." He will supplement these sources with his own fund of informal intelligence gleaned from experience.

Second, the reference specialist must provide the professional analytical capability to undertake system-oriented research of user requirements and develop on a time phased basis the operating characteristics of the reference system.

Third, the reference specialists must undertake the task of constructing the reference system and its component formal reference index. This will require bringing together and systematizing large amounts of technical intelligence and incorporating it into an operating system.

One of the biggest problems in developing a reference capability is that the kind of professional reference specialist that is characterized here does not exist. It is a new kind of professional capability of emerging importance. There are a few men in established agencies whose work experience fits them with attributes that come close to the functional requirements outlined above. They are very limited in number, however, and, characteristically, are key men in fulfilling agency missions.

The success of the attempt to develop a reference system will rise or fall on the strength of the kind of professional talent that will guide its development. Since the kind of experience that is necessary in this function is rare, an essential part of any program effort will be an explicit recognition of this fact and an explicit procedure for the training and development of statistical reference specialists.

Ways will have to be worked out so that new professionals could have rotating assignments that would carry them into the primary statistical agencies where they could (1) work on specific components of reference information for the purpose of implementing the reference file, (2) come under the supervision of and receive training from those people who represent the greatest fund of accumulated knowledge, and (3) receive a total system orientation that could be gained in no other way. Project financing would have to incorporate explicitly the resources that would support staff training and development.

The task is made somewhat less formidable by the fact that the three categories of reference functions outlined above are susceptible to some degree of specialization. During developmental phases the intellectual resources of the staff could also be supplemented through consulting arrangements. It would be useful, for example, for the operating agencies to assign some of their specialists to work with the reference service on a temporary or part-time basis. The analytic or system design component might be especially amenable to supplementation during the early phases.

##### *The organizational form*

The feeling was strong that a successful effort to develop a Statistical Reference Service would require some degree of centralization of function. An important part of the reference function is interagency or total set in character and cannot be handled within the context of an agency orientation. At the same time, the reference agency can perform its function without involving the primary agencies directly. In the previous section we already outlined some of the ways in which agency participation would be indispensable. The agencies will need to play a role as a breeding ground for research specialists, and, of course, will be the source of most of the reference intelligence that must be used to construct and maintain an index. It may also be necessary and desirable to establish within at least some of the larger agencies a companion reference function that would be linked with the total reference capability.

The committee did not discuss the specific institutional form of such a service center within the framework of the existing institutions.

#### COSTS

The committee devoted some time to a consideration of the resources that would be required to establish such a reference service. It came to the conclusion that nothing very precise could be said about costs at this stage for several reasons.

First, the kind of evaluation that can be given by a group of this type at an early speculative state is suspect. A more refined notion would require the application of more staff resources to planning and evaluation than are currently available.



Second, the question of costs is confused by the possible existence of joint costs in this program area. The general problem set of which the reference problem forms a part has several other dimensions that extend beyond the restricted problem this committee has taken for discussion. The costs of establishing a reference service of the kind discussed here would be quite different if it were established as a discrete function of, if it were developed as one component of, a more generalized user service capability. If the provision of reference services were combined with other statistical services addressed to establishing and maintaining archives, servicing administrative records, or providing tape translation, disclosure bypasses and other file management services, the professional staff and program facilities including computer facilities could perform many joint functions. The increment costs of providing a reference capability as a part of a more extensive user service capability could possibly be a great deal smaller than would be required for a separate and independent function.

Despite these obstacles to cost estimation, the committee noted that the operating budgets for two agencies performing related (though in many ways basically different) functions ranged from \$400,000 a year for the National Science Referral Center in the Library of Congress to \$2 million a year currently for the Science Information Exchange of the Smithsonian Institute. The committee speculated that it probably would take as much as \$2 million a year on the average over a 5-year period to develop a meaningful reference service capability including enough resources to involve agency programs in the way necessary.

The committee also discussed the question of the demand for reference services and whether anticipated needs were appropriate to justify such levels of expenditure. The need for the service was judged to be sufficiently great to justify a serious effort.

There is no way, at this preliminary stage, that anything more than a judgment can be offered. In the first place, nothing in present agency experience can serve as a guide to demand levels for a service that has never existed in anything like the form indicated in this prospectus. The judgment rested on several considerations. First, there is considerable evidence of frustration and inefficiency because of the absence of such a service. Second, there already exist other programs, such as the ones referred to, that have been judged essential and for which expenditures substantially exceeding the sums of money mentioned here have been undertaken. Most of these have been undertaken in the interest of improving the efficiency of documentation in the physical sciences. In the circles where public and business policy are made, and social science, management and marketing research undertaken, the most compelling needs is not for a way to handle better the traditional documentary materials generated, but for a way to acquire efficient access to relevant numerical files that constitute the main bodies of evidence and of research inputs. The need for expanding the services in this area seems equally as compelling as those in the field of physical science documentation that are already receiving extensive attention.

## APPENDIX B

OCTOBER 1, 1965.

From: Rudolph C. Mendelssohn.  
To: Mr. Edgar Dunn.  
Subject: Report on data inventory.

## DATA BANK REQUIREMENTS

An estimated \$3 to \$3.5 million and 3 to 5 years are needed to stock the proposed data center with Federal statistics now in existence. These data would probably comprise about 20,000 reels of magnetic tape. However, a Federal center could be stocked with a respectable volume and variety of data relatively quickly for about \$260,000. At the rate of about \$27 per reel, a bank of 9,000 tape files could be established in about a year. Such a course would provide a fairly representative selection of significant data including, for example, 750 reels of the census housing data; census current population data on 375 reels; the BLS Consumer Expenditure Survey on 43 reels; the BLS industry hours, earnings, and labor turnover data on 36 reels; the OBE national income accounts on 2 reels; as well as IRS tax data on 5,300 reels and BOASI social security data on 1,900 reels. An insight into the volume of data readily available is gained by the rough calculation that the equivalent of nearly 1 billion punched cards would be included (table 1).

Of the \$3 to \$3.5 million needed to supply the data center over the 5-year period, about \$700,000 would be needed to bring data not now machine accessible into usable form. About \$500,000 would be needed to transfer punched card data to magnetic tape. And about \$1.5 million would be needed to reprocess data now on magnetic tape. Looked at another way, about 100 man-years of professional aid will be required to review and correct the records, develop computer systems and programs to process the data, and provide reference documents to show file contents to the serious scholar from both the substantive and machine processing points of view. Also, about 100 man-years of keypunching and a very rough estimate of 10,000 computer hours are indicated (table 2). In contrast, about 13 man-years of professional work and 1,200 hours of machine time would stock the center with half the total in a year.

The major resource requirement of the 5-year effort is for the Census Bureau where \$1¼ million is requested. Over half the amount reported for that Bureau, about \$700,000, is needed to bring the 25- and 5-percent population samples for 1960 to acceptable levels. On the other hand, the Internal Revenue Service and the Bureau of Old Age and Survivors Insurance, both among the giants of data processing, have requested relatively limited amounts \$17,400 and \$14,300, respectively.

In considering the course of stocking the center several key factors should be kept in mind. First, the cost of additional historical reels after the initial storage of 9,000 is quite high—about \$300 per reel, compared with about \$27. The cost per reel is considerably higher for some of the files. For example, the SEC Quarterly Financial data and the FRB Report of Condition of Insured Banks would cost \$20,000 and \$8,000 per reel, respectively. Obviously, a careful review and justification for high-cost files is needed before their improvement can be supported. Second, the costs indicated in this report refer only to those needed to make data accessible within the responsible agency. I am assuming that the proposed data center would defray the costs of tape copying and would supply its own blank reels. Such costs are not inconsiderable. The 9,000 reels which could be made ready in about a year would cost the center over a half million dollars for blank reels and for copying.

## GENERAL COMMENTS

I have the general impression that the larger the volume of data and the higher the frequency of processing the greater is the tendency for the files to be in acceptable order. That is to say, large files like those of the IRS and BOASI and the high-frequency operations in the BLS manpower field and in the FRB are in good shape while some decennial and annual operations at the Census Bureau and the relatively small files at SEC, FTC, OBE, and the Office of Education are either not well maintained with the computer or are not well mechanized at all. In other words, the degree of accommodation to the computer seems to be a function of the work pressures to use it.

The vast majority of available information is already in machine form. The small amounts of significant data not machinable are found in OBE, the Office of Education, and in the Department of Agriculture. The files in OBE and OE are not large, would total about 200 reels, in my estimation, and are not in machine form in appreciable amounts. As you know, efforts are underway to correct this in both agencies. About 25 percent of the Agriculture data are now machinable, according to Department representatives. It was asserted that the remaining 75 percent are significant and useful data and should be available to a data bank for research in agricultural economics.

## AGENCY COMMENTS

Three general comments tended to be made by agency representatives. First, and least frequent, assertions that the Budget Bureau survey and the Ruggles Committee report has brought an increased awareness of the need for more effective file maintenance and that efforts to achieve this end would be incorporated in ongoing operations where feasible. These good intentions should be supported with funds where appropriate and the dilemma in which the Census Bureau finds itself with respect to the 1960 population samples should not be allowed to be repeated. I have no doubt that the \$700,000 now required would have been far less if the job had been done as a part of the 1960 census work.

Once the records have been brought to acceptable levels through new financial support, I doubt that programs operated at high frequencies will require more

than trifling amounts to maintain this level in current operations. On the other hand, many of the recurring annual and decennial operations need explicit support for the improvement of file management practices. The 1960 population samples are examples of the latter situation. An assessment of these costs is outside the scope of this survey and, in fact, should be considered a part of the cost of the survey without regard to a proposed Federal data centers.

Then, there was the comment by the larger agencies, heavily stressed by the Census Bureau, that funds to improve the files would not necessarily generate the desired results. The argument was that only present staff members had the background and professional experience needed to do the job. Since these people were already fully occupied, could not be diverted, and new staff could not do the job, it would not be done even if money is supplied it was asserted. I do not concur with this view. The work can be done if it is properly phased; that is, if sufficient time is allowed, new workers under the supervision of experienced personnel can do the job in the long run (the 5-year span I have suggested). I believe the reaction cited above assumes a crash effort to organize the files on a high-priority basis. I agree that it cannot be done this way and advise against such an approach.

Agency representatives seemed excessively concerned with the confidentiality question. Turning data over to a Federal center would be a breach of contract with respondents who have been assured that none but agency personnel would view their reports, it was said. I tried to convey the assurance that, if a data center were established, it would assume the obligation of protecting both the agency and the respondent. Since feelings on this matter run quite deep, some steps should be taken at the outset to vitiate them or discussions beyond this narrow consideration could founder.

One constructive suggestion was made in regard to confidentiality. Mr. Robert Menke of the Securities and Exchange Commission expressed the view that corporate concern dealt mainly with current affairs. It was his feeling that, after a period of 5 to 10 years, back data could be exposed to public view without serious objection by respondents. There would be difficulty perhaps in applying such a rule retroactively but a notice to this effect on future collections of data might serve to make the problem less troublesome in the years ahead.

#### CONCLUSION

I have a final comment. I found the evident loss of data because of the failure to support good file management distressing. Immediate steps ought to be taken to stop this erosion of a national resource. The costs indicated above measure the deficiency of not doing it before, and they will grow as time passes. It is difficult to argue that these losses have immediate meaning. The tools, techniques, and intellectual attitudes needed for their useful exploitation are not yet reflected in our institutions. But, as you know, changes are already underway. And even though we are unable now to predict how the store of data might be used. I am convinced that actions to preserve this national resources will be appreciated by those who follow.

TABLE 1.—Data file inventory—Selected files which can be ready in 1 year, by agency and project

File name and data type	Current number of tape reels (1)	(2) Improvement through added resources				Cost of documentation (3)	Total costs (4)
		Substantive personnel (a)	Systems analysis and programming (b)	Machine processing (c)	Number of tape reels after additional resources (d)		
Total, all agencies.....	10,340	\$11,100	\$71,540	\$126,125	8,971	\$52,710	\$261,475
Census Bureau.....	2,230	11,100	41,600	78,500	1,407	24,550	155,750
County City Data Book, 1952, 1955, 1962.....	3				3		
1960 Census of Housing, basic 25 percent summaries.....	1,474		12,000	55,000	750	7,500	74,500
Quarterly Survey of Intentions.....	1,172		5,000	6,000	172	1,000	12,000
Monthly Current Population Survey.....	375		22,000	8,000	375	15,000	45,000
Annual Survey of Manufacturers, 1959-62.....	64	8,000	1,100	5,500	40	100	14,500
Current retail report.....	32	100	1,500	4,000	32	100	800
County business patterns.....	100	3,000			25	200	8,700
Manufacturers shipments, inventories, orders.....	10				10	250	250
Bureau of Labor Statistics.....	134		23,000	20,500	79	2,300	54,700
Survey of Employment, Payroll, and Hours.....	84		23,000	25,000	30	400	48,400
Survey of Industry Labor Turnover.....	2				2	400	400
Survey of Employment, Payroll, and Hours.....	1				1	400	400
Consumer Expenditure Survey, 1960-61.....	44			3,500	43	600	4,100
Office of Business Economics.....				1,000	2		1,000
National Income and Production Accounts.....				1,000	2		1,000
Internal Revenue Service.....	5,303				5,303	17,400	17,400
Individual income tax returns.....	2,503				2,503	3,000	3,000
Corporation income tax returns.....	1,475				1,475	4,500	4,500
U.S. business tax returns, sole proprietorships.....	764				764	3,000	3,000
U.S. partnerships returns.....	232				232	2,050	2,050
Tax model for individual income tax returns.....	316				316	2,850	2,850
Tax model for corporation income tax returns.....	20				20	1,000	1,000
Tax model for partnerships returns.....	3				3	1,000	1,000

TABLE 1.—Data file inventory—Selected files which can be ready in 1 year, by agency and project—Continued

File name and data type	(2) Improvement through added resources					Cost of documentation (3)	Total costs (4)
	Current number of tape reels (1)	Substantive personnel (a)	Systems analysis and programming (b)	Machine processing (c)	Number of tape reels after additional resources (d)		
Total, all agencies—Continued							
Federal Reserve Board							
1963 Survey of Financial Characteristics	74				50	\$150	\$150
Reports of condition of all insured commercial banks	25				1		
Reports of income and dividends, all insured banks	35				35		
Daily money supply	11				11		
Industrial production index	2				1	150	150
National Center for Health Statistics							
Annual questionnaire	220		\$4,200	\$10,125	215	5,050	19,375
National Vitality Sample Survey, 1963	35		400	1,000	35	1,800	3,200
National Mortality Sample Survey, 1961-65	2		400	25	1	200	625
National Mortality Sample Survey, 1960	4		500	75	1	350	925
Annual birth and fetal death statistics	2		400	25	1	200	625
Annual death statistics—detailed data	87		1,000	4,000	87	1,000	6,000
Bureau of Old Age and Survivors Insurance							
1 percent continuous work history sample, 1937 to date	2,379		2,740	8,000	1,915	3,380	14,100
0.1 percent continuous work history sample, 1937 to date	196		870	4,000	120	1,080	5,950
Name and address file tape	39		870	1,000	31	1,080	2,950
Summary earnings tape	94				94		
Regular transcript master benefit tape	1,600		1,000	3,000	1,600	200	200
	450				70	1,000	5,000

TABLE 2.—Partial data file inventory—Totals by agency and type of data

Agency	Current number of tape reels (1)	Professional (substantive, system, and documentation) costs (2)	Transcription costs (3)	Machine processing costs (4)	Total costs (5)	Number of tape reels after additional resources (6)
Total, all agencies	18,831	\$1,239,610	\$544,300	\$1,017,825	\$2,788,235	16,269
Source	18,338	999,410	503,000	939,440	2,441,850	15,988
Summary	493	240,200	41,300	78,385	346,385	282
Bureau of Labor Statistics:						
Total	385	218,100	32,000	81,400	318,000	124
Source	148	68,900		44,100	113,000	102
Summary	237	149,200	32,000	37,300	205,000	22
Bureau of Employment Security:						
Total	32	6,900	1,000	22,650	30,550	74
Source						
Summary	32	400	1,000	850	2,250	1
		6,500		21,800	28,300	73
Bureau of the Census:						
Total	10,047	566,550		715,050	1,281,600	8,106
Source	9,838	545,800		705,050	1,250,850	7,994
Summary	209	20,750		10,000	30,750	112
Office of Business Economics:						
Total	1	10,900	300	2,250	13,450	5
Source		6,000		1,500	7,500	2
Summary	1	4,900	300	750	5,950	3
Department of Agriculture:						
Total	155	119,500		37,290	156,790	151
Source	155	91,500		35,780	127,280	141
Summary	0	28,000		1,510	29,510	10
Internal Revenue Service:						
Total	5,303	17,400			17,400	5,303
Source	5,303	17,400			17,400	5,303
Summary						
Federal Trade Commission:						
Total	0	8,000	2,000	3,000	13,000	7
Source	0	7,000		2,000	9,000	6
Summary	0	1,000	2,000	1,000	4,000	1
Office of Education:						
Total	10	2,315	0	2,570	4,885	21
Source	0	2,115	0	2,370	4,485	10
Summary	10	200	N.A.	200	400	11
Bureau of Old Age and Survivors Insurance:						
Total	2,538	32,650		110,500	143,150	2,074
Source	2,538	32,650		110,500	143,150	2,074
Summary						
Securities and Exchange Commission:						
Total	0	213,500	56,500	9,400	279,400	12
Source	0	193,500	52,000	9,250	254,750	11
Summary	0	20,000	4,500	150	24,650	1
Federal Reserve Board:						
Total	75	30,050	452,500	15,640	498,190	125
Source	71	20,400	450,000	15,090	485,490	109
Summary	4	9,650	2,500	550	12,700	16
National Center for Health Statistics:						
Total	285	13,745	0	18,075	31,820	268
Source	285	13,745	0	12,950	26,695	235
Summary	0	0	0	5,125	5,125	33

TABLE 2, PART I.—Partial data file inventory—Data on magnetic tape, by agency and type of data

Agency	Current number of tape reels	Improvement through added resources				Cost of documentation	Total costs
		(2)					
		Substantive personnel costs	Systems analysis and programming costs	Machine processing costs	Number of tape reels after additional resources		
	(1)	(a)	(b)	(c)	(d)	(3)	(4)
Total, all agencies	18,831	\$211,200	\$304,915	\$840,880	15,717	\$126,680	\$1,484,675
Source	18,338	172,300	267,615	814,830	15,545	110,380	1,365,125
Summary	493	38,900	37,300	27,050	172	16,300	119,550
Bureau of Labor Statistics:							
Total	385	29,400	59,700	46,100	101	20,000	155,200
Source	148	2,900	26,100	31,600	86	10,400	71,000
Summary	237	26,500	33,600	14,500	15	9,600	84,200
Bureau of Employment Security:							
Total	32	1,000	1,000	3,000	32		5,000
Summary	32	1,000	1,000	3,000	32		5,000
Bureau of the Census:							
Total	10,047	176,100	201,600	658,500	7,810	66,850	1,103,050
Source	9,838	165,000	199,000	649,000	7,700	65,800	1,078,800
Summary	209	11,100	2,600	9,500	110	1,050	24,250
Office of Business Economics:							
Total	1	300	100	50	1	500	950
Summary	1	300	100	50	1	500	950
Department of Agriculture:							
Total	155	4,400	7,500	10,780	100	5,400	28,080
Source	155	4,400	7,500	10,780	100	5,400	28,080
Internal Revenue Service:							
Total	5,303				5,303	17,400	17,400
Source	5,303				5,303	17,400	17,400
Office of Education:							
Total	10	(1)	(1)	(1)	10	(1)	(1)
Summary	10	(1)	(1)	(1)	10	(1)	(1)
Bureau of Old Age and Survivors Insurance:							
Total	2,538		29,290	110,500	2,074	3,360	143,150
Source	2,538		29,290	110,500	2,074	3,360	143,150
Federal Reserve Board:							
Total	75				51	5,150	5,150
Source	71				47		
Summary	4				4	5,150	5,150
National Center for Health Statistics:							
Total	285		5,725	12,950	235	8,020	26,695
Source	285		5,725	12,950	235	8,020	26,695

1 Not available.

TABLE 2, PART II.—Partial data file inventory—Data on punched cards, by agency and type of data

Agency	Cost of professional services	Machine processing costs	Tape reel equivalents	Total costs
	(1)	(2)	(3)	(4)
Total, all agencies	\$367,415	\$139,120	478	\$506,535
Source	292,715	106,285	376	399,000
Summary	74,700	32,835	102	107,535
Bureau of Labor Statistics, total	64,000	18,800	21	82,800
Source	29,500	12,500	16	42,000
Summary	34,500	6,300	5	40,800
Bureau of Employment Security, total	4,500	18,800	41	23,300
Summary	4,500	18,800	41	23,300
Bureau of the Census, total	122,000	56,550	296	178,550
Source	116,000	56,050	294	172,050
Summary	6,000	500	2	6,500
Office of Business Economics, total	6,000	1,500	2	7,500
Source	6,000	1,500	2	7,500
Department of Agriculture, total	102,200	26,510	51	128,710
Source	74,200	25,000	41	99,200
Summary	28,000	1,510	10	29,510
Federal Trade Commission, total	7,000	2,000	6	9,000
Source	7,000	2,000	6	9,000
Office of Education, total	2,315	2,570	11	4,885
Source	2,115	2,370	10	4,485
Summary	200	200	1	400
Securities and Exchange Commission, total	57,500	6,775	5	64,275
Source	57,500	6,775	5	64,275
Federal Reserve Board, total	1,900	490	12	2,390
Source	400	90	2	490
Summary	1,500	400	10	1,900
National Center for Health Statistics, total	0	5,125	33	5,125
Summary	0	5,125	33	5,125

TABLE 2, PART III.—*Partial data file inventory—Data not on machineable forms, by agency and type of data*

Agency	Professional services cost	Transportation costs	Machine processing costs	Tape reel equivalents	Total costs
	(1)	(2)	(3)	(4)	(5)
Total, all agencies.....	\$229,400	\$544,300	\$23,325	75	\$797,025
Source.....	156,400	503,000	18,325	67	677,725
Summary.....	73,000	41,300	5,000	8	119,300
Bureau of Labor Statistics, total.....	45,000	32,000	3,000	2	80,000
Summary.....	45,000	32,000	3,000	2	80,000
Bureau of Employment Security, total.....	400	1,000	850	1	2,250
Source.....	400	1,000	850	1	2,250
Office of Business Economics, total.....	4,000	300	700	2	5,000
Summary.....	4,000	300	700	2	5,000
Federal Trade Commission, total.....	1,000	2,000	1,000	1	4,000
Summary.....	1,000	2,000	1,000	1	4,000
Securities and Exchange Commission, total.....	156,000	56,500	2,625	7	215,125
Source.....	136,000	52,000	2,475	6	190,475
Summary.....	20,000	4,500	150	1	24,650
Federal Reserve Board, total.....	23,000	452,500	15,150	62	490,650
Source.....	20,000	450,000	15,000	60	485,000
Summary.....	3,000	2,500	150	2	5,650

## APPENDIX C

## THE DESIGN OF A FEDERAL STATISTICAL DATA CENTER

(A report to the Bureau of the Budget, prepared by E. Glaser, D. Rosenblatt, M. K. Wood, National Bureau of Standards)

## SUMMARY AND CONCLUSIONS

This report was prepared in response to a letter from R. T. Bowman, Assistant Director for Statistical Standards, Bureau of the Budget, to A. V. Astin, Director of the National Bureau of Standards. The original request was focused principally upon "the possibilities for developing new capabilities in computer hardware systems that will improve their flexibility and economy in specialized file storage, management and retrieval functions" in connection with a national data service center. It directly became evident that any such study would be substantially conditioned by the characteristics of the information system to be mechanized: the scope and content of the economic and demographic data to be included; the degree of detail for each kind of information; the ability to use (a) data of more than one kind in a single analysis or mathematical model, (b) data derived from different reporting systems, (c) data collected by different agencies, (d) data referring to different time periods; the kinds of access to the files and the formats of acceptable queries; the nature and extent of computational and manipulative services to be provided; and other system specifications. Since there was no definite specification with regard to these characteristics, it was decided that a preliminary description of such a system was prerequisite to the requested analysis of hardware characteristics.

Mr. Edgar S. Dunn, consultant to the Office of Statistical Standards, Bureau of the Budget, worked with the staff of the National Bureau of Standards in reinterpreting the original request. As a result, the group's efforts were directed toward the issues that would govern the design of an effective Federal Statistical

Data Center. The original intent to explore the potentialities of modern large-scale computers is constantly in evidence in the present formulation.

In preparing the report, the problem of describing the customer population was considered first. Something is known of the kinds of specialists who use data originating in the Federal Government to solve problems in economic policy, public administration, business economics, business administration, and a great range of social science subjects. But it is also apparent that this present user population reflects the capabilities and logistics of present ways of organizing and purveying data. At least equal importance attaches to those needs which are not met by present practices. How can these unmet needs be characterized?

An adequately specified information system would have to be based upon a broad review of the types of analyses that a wide spectrum of social scientists propose and upon the quantitative models that they build. For the present purposes and the limited scale of effort, it was decided to restrict the review to several classes of economic models directed at problems of national economic policy. Even this limited review revealed a variety of possible requirements for socioeconomic information which are not now being met, although many of the basic data are collected and compiled in some form by some Federal agency.

The review of economic models and their needs for statistical information was conducted at a 4-day conference at Fort Ritchie, Md. on August 26-29, 1965. Participants were M. K. Wood, D. Rosenblatt, and E. Glaser of the National Bureau of Standards and E. S. Dunn and P. F. Krueger of the Bureau of the Budget.

Subsequent conferences and staff work built upon the Fort Ritchie conference by developing (a) an enumeration of the services to be rendered, and (b) a description of the Federal Statistical Data Center in terms of its functions and principal characteristics. A summary of these is given below.

A. *Nature of the services to be rendered*

An integrated Federal Statistical Data Center appears necessary to perform the following functions:

1. To provide data in cases where the primary agency in possession of the data is not capable of making it available in the required format, detail, flexibility, or quality.
2. To provide data where the information originates in two or more reporting systems or agencies, in order to make available information about interrelationships in maximum feasible detail, without restrictions resulting from screening for improper disclosures at the time of transfer into the Center and through association of information from multiple sources relating to the same individual reporting unit or analytical unit.
3. To maintain an archive of statistical data, complete in the sense described in 2 above, with all corrections and adjustments carried through in a consistent manner, and with a collection of the accompanying codebooks and manuals.
4. To provide information outputs (responses to queries) in a variety of forms at the customer's option: printed tabulations; machine readable tapes, graphs, diagrams, etc., either locally or through telecommunications.
5. To establish, maintain currently, and operate a reference and referral service for the Federal statistical system.

The creation of such a Federal Statistical Data Center also should provide the following additional services an corollary benefits at minimum cost:

6. ADP equipment would be available for computation and data reduction in response to queries of customers: cross tabulations, averages, distribution statistics, smoothed curves, trend fittings, seasonal adjustments, periodic analyses, correlations, regressions, and more advanced analyses in order to give access to the full range of information computable from the collection.
7. Confidentiality audits would be performed by machine upon the information intended for release to customers.
8. ADP equipment would also support a battery of services to the statistical system of the Federal Government: computations essential to the conduct of test adjustments on statistical series and collections, computations for test reconciliations of data from two or more sources or for two or more time periods, detection of errors in primary collections or derived statistics through consistency tests and anomaly detection routines, computations necessary for the study of error propagation through the Federal statistical system, combinations of the above computations in support of validation studies for Federal statistics and in support of procedures for certification of the accuracy and consistency of Federal statistics.



9. ADP equipment would also be used to provide service agencies with large-scale adjustment and reconciliation tasks (which is already being done by some agencies) in the production of standard series and to assist in the creation of new series through the reconciliation, adjustment and transformation of standard series.

*B. Character and organization of the data in a Federal Statistical Data Center*

1. Principles must be developed for the initial selection and future accession of data to be included in the Center's collection; they should reflect a broad range of uses and full utilization of basic information rather than a codification of present uses, present practices, and present compromises.

2. Methodology and principles must be developed for the conversion of present files and production data to suitable archive form and quality, and their maintenance in proper form and quality, supported by codebooks, manuals, etc.

(a) The principles established for archives must presume that data will be transferred from the collecting agency to the Center in full available detail.

(b) The principles established for archives must provide for the systematic completion of all corrections and adjustments to all data affected and all levels of detail, resulting in a fully reconciled and consistent body of data.

(c) The principles established for archives must provide for data to be transferred to the Center without screening for confidentiality; all confidentiality audits and checks would be applied to the formats and information content of the output of the Center.

(d) The principles established for archives must provide for the preservation of the identity of the reporting unit and the association of all information about the unit without regard to the agency or manner of the collection of the information; the rules and the economics of matching of existing records will be very different from those intended for future production of data.

3. Standards must be developed for definition, coding, classification and aggregation with the intent of maximizing the ability to use different kinds of data in the same analysis and of minimizing the loss of information.

4. Standards must be developed for formats in order to facilitate the management, housekeeping and retrieval of records and to avoid the loss of information.

5. Standards must be developed for quality of data. (consistency of definition, error rates, etc.) and for means of assuring maintenance of quality.

6. Automatic data processing (ADP) equipment and systems must be available to the Center to provide economy, timeliness, and flexibility of access to the information in the records.

(a) ADP must be available

(1) to compute statistics that are inherently computable from the records,

(2) To provide answers in the required form, and

(3) To avoid unnecessary withholding of information as a result of using inefficient and redundant disclosure criteria.

(b) ADP must be available for the conduct of confidentiality and other disclosure audits, such rules and procedures to be applied to the data in the form and content intended for release from the Center.

7. Criteria must be developed for assuring that the Center is established, and is maintained, in a manner that is responsive to a broad base of potential users, rather than in a manner which seems to suit the present habitual users at any time.

8. Criteria must be devised for periodic review of the value of the data contained in the archive followed by a selective purging of the data whose retention is no longer justified.

DESCRIPTION OF A FEDERAL STATISTICAL DATA CENTER

*Introduction—Federal statistics and the computer*

The statistical services of the Federal Government were initially created in response to a variety of unrelated needs. The census of population, in its early simple form, arose in response to a specific constitutional provision. The first census of manufacturers was a purposeful study of the existing status and likely potential economic development of the young Nation. Other collections of data were directed at a continuing study of the operation of the economy: prices,

employment, sales and inventories, production of specific minerals, etc. A large number of statistical collections arose in connection with specific Federal programs, the statistics themselves being largely byproducts: activities and finances of regulated industries, internal revenue statistics, health and educational defects among selective service registrants, grants to scientific researchers, veterans' benefits, etc.

Two major influences have been at work to give improved quality and cohesiveness to this initially piecemeal collection of information about all aspects of the Nation and its people. One was primarily organizational and the other technical.

The Federal Reports Act of 1942, building upon such earlier programs as that of the Central Statistical Board of the National Recovery Administration, created a coordinating mechanism for the improvement and rationalization of the Federal statistical system. The Director of the Bureau of the Budget was given staff and responsibility for the development and introduction of standards for the collection, processing, and dissemination of data through much of the Government's activities. The standard industrial classification, the standard metropolitan statistical areas, the standard sample week for monthly surveys, and the standard base periods for economic time series are examples of this standards-setting function. The financial reporting program and the current population and labor force program are examples of the coordination of the work of several agencies to produce data useful for a variety of purposes. The Bureau of the Budget had become an instrument for promoting systematic cooperative efforts among the many Federal, State, and private collectors and processors of information.

The technical base for improved quality of information also has roots in the past. The development of improved techniques for acquiring information, analyzing it, preparing it for publications, and using it for economic, social, and political studies has been active for over a century. With the growth of applied social sciences and the progressive elaboration of the Government's statistical activities, the pace of research in technical methods was greatly accelerated. Trained statisticians brought improved techniques to many aspects of their work: sampling, the design of experiments, seasonal adjustment of time series, the construction of national and regional accounting models, the study of non-sampling errors in surveys and censuses, the development of quality control and other sequential methods, and the interpretation of data in complex situations. The consequences of these technical improvements have been far reaching.

During and directly after World War II, the design and construction of the first electronic computers foretold a potential for vast improvements in many aspects of statistical technology: the recording and editing of field survey data; the compiling, tabulating, and publication of data; the analysis of data and their use in problem solving. The Federal Government pioneered in exploiting these new capabilities. However, much more can be done, particularly in the design of better ways of organizing economic and social data, more thorough integration of information from the many separate statistical programs, and the reduced loss of information in utilizing data for analytical purposes and purveying it to various classes of customers.

The very general logical powers, the great storage capacity, the high speed of manipulation, and the low unit cost of modern ADP systems combine to promise great potential improvement in information resources and problem-solving capabilities. A number of Federal agencies have learned the advantages of mechanizing their routines. Indeed, the Bureau of the Census contracted for the development of the UNIVAC, the first commercial internally programed computer, and it acquired the first and fourth units produced. Many of the frequently cited Federal statistics are more promptly and more satisfactorily produced than would be possible without computers; whether the statistics arise from a primary function of the agency or as a byproduct.

Yet, the improvements were typically made within the context of a single agency—usually a single reporting system—and without the possibility of raising broad questions about the fundamental organization of the Federal statistical system as a potentially unified and cohesive collection of intelligence. Nor was there any practical way of applying the explosively growing power of computers to general questions of preventing loss of information once it had been brought into the system by one or another Federal agency. Finally, there has been no serious attempt to assess the consequences of the computer for improved access to Federal data or for meeting the need for providing information in the form, degree of summarization, format, and physical output desired by various classes of customers.

The present report is addressed to these questions

*General principles for the organization of data for a Federal Statistical Data Center*

The consequent reconsideration of the organization of socio-economic data in the Federal Government is based upon two general guides. The first is to review the implicit informational requirements of the whole range of analyses and formal models proposed by social scientists, rather than to restrict the statement of information requirements to those needs which have been given principal attention in the past. The second is to consider the whole range of relevant tasks that the computer can assist, even if the manner of proceeding is radically different from current practice.

From this reconsideration four general principles emerge for constructing specifications for a Federal Statistical Data Center:

1. Maximum ability to exhibit the interrelations among various kinds of data;
2. The unification of all information about the individual reporting unit or analytical unit;
3. The preservation of detail in the basic records and the avoidance of loss of information in the storage, manipulation, and retrieval of information; and
4. The ability to produce the full measure of inherent information which is computable from the basic records.

These four principles will now be developed as groundwork for specifications for the information organization and the services of a Federal Statistical Data Center.

One of the greatest deficiencies of the existing Federal statistical system is its failure to provide access to data in a way which permits identification and measurement of functional interrelationships among interdependent activities. Identification and measurement of such interrelationships are essential to a wide range of economic and social analyses. It is also the chief problem in the design of mathematical models of economic and social processes suitable for appraising the impact of alternative policies and programs as well as possible changes in environmental factors.

Such appraisal is, in turn, prerequisite to effective benefit-cost analysis of proposed and ongoing programs. The essence of rational benefit-cost analysis is the tracing of indirect as well as direct effects of programs and the evaluation and summing of these effects. Typically, the methodology for tracing all but the most obvious linkages is entirely lacking or fails to use the relevant information.

Until recently, economic model builders have been restricted to relatively aggregative economic and resource flow models, and to inferring interrelationships among very few aggregative variables. Such relationships often have considerable predictive value where other conditions remain relatively stable or continue to change at a constant rate. But the essential relationships are correlative or associative rather than structural. Hence, they generally fail to give acceptable prediction when other conditions change markedly, as a result of changes in major program, policy, or environmental factors.

Acceptable prediction under changing circumstances requires analytical models which give much more detailed and explicit recognition to interrelationships among the criteria and variables which will be affected by the changed conditions. Such analytical models generally describe the mechanisms in greater detail than the associative models; they use more information, and they often rely less heavily on trends or the postulation of only slow changes among the variables in the model. The present and prospective accelerated pace of technological and statistical change now requires the development and use of more detailed and complex models than can be created or supported by the present Federal statistical system.

The rapidly developing tools of automatic data processing and systems analysis now make possible—and necessary—both the development of more advanced models and the elaboration of the Federal statistical system which is needed to support them.

Many of the data needed for establishing causal interrelationships among related economic variables are contained in the existing Federal statistical system. But present collection methods, tabulation procedures, and disclosure rules combine to make it difficult and often impossible to extract such data. Where samples are largely enough, it may be possible to cross-tabulate in a way which permits determining the interrelationships between two variables or, rarely,

among three. But generally when more than a two-way relationship is involved, it is impossible to tabulate necessary totals in a way which will define the desired relationship without disclosure of proprietary data. Such relationships among many variables can be extended, however, at a much greater level of detail, if it is possible to apply standard statistical analysis techniques to the observations for individual respondent units over the whole range of the relevant variables. It is possible in this way to extract much more useful structural information and still insure that no disclosure of individual respondent data is contained in the results of such analyses.

In many cases, the data necessary to such an analysis require the matching of items from two or more statistical sources. One important class of analyses involves the matching or reports by the same respondent for different time periods. Some data files are so organized as to make this possible, but many are not. An even more complex problem arises when it is necessary to match data from the same respondent collected as parts of different statistical programs, by different agencies. This can be extremely difficult or impossible, though substantial progress has been made in some areas, as, for example among Census, Bureau of Old Age and Survivors Insurance (BOAST), and the Internal Revenue Service. Problems of disclosure are most difficult in this context.

There are several fundamental problems dealing with the coding and classification of original source data. Most serious is the need for uniform identification, definition, and coding of the respondent unit. Unless this is done, matching of data from diverse sources is generally impracticable if not impossible. A uniform system of classification and coding for geographic area is another major deficiency.

In general, the classification and grouping of data are dictated by the problem environment, the basic logic of the analytical model, and the kind and degree of detail in which the results must be expressed and interpreted. In practice, there is frequently need to compromise the ideal classifications and aggregations of data for several reasons; the basis and criteria of classification in the collection agency being inconsistent with the ideal requirements of the model; the lack of sufficient detail (industry, process, product, geographic location, etc.); the withholding of detail under proprietary confidentiality or security restrictions; the noncompatibility of the definitions of the respondent units in the several collection systems which could otherwise provide the information specified by the model, which can be reconciled only by coarse aggregation but with accompanying loss of information and structural detail; the noncompatibility of classification of the data by several collection agencies and information systems also capable of specious resolution by aggregation; the difficulty and cost of identifying and matching the reporting units from two or more reporting systems, so that the information about the reporting unit can be pooled; the absence of technique, staff, funds, and machine time to use large-scale data processing equipment to recode, recompile, reconcile, reclassify, and aggregate data and to perform all manner of statistical procedures upon the data.

Since there are very large numbers of ways in which most economic variables might reasonably be classified and aggregated, it is not practical to prepare the data in all of these formats in anticipation of possible requests. Nor is this necessary. The same results can be achieved with favorable logistics and great flexibility by providing for the basic records to be maintained in machine-receivable form and in as fine detail of classification as is practical. The low unit cost and high speed of modern computers can then be exploited to meet requests for data with little loss of the available information inherent in the combined resources of the participating agencies.

The availability of modern computers can meet two important requirements in this context. The first is discussed above: the conversion from finely disaggregated classes to all manner of special purpose classifications and aggregations (and, indeed, conversion to publishable forms). The second requirement is to avoid unnecessary loss of information because of proprietary and confidentiality restrictions. The fundamental rule in this case is to perform all edits and checks relating to unwanted disclosure upon the fully processed data (aggregations, summaries, averages, correlation coefficients, regressions, fitted curves, etc.) rather than upon the detailed raw data. This will assure full use of information consistent with disclosure rules. The logical capability of the computer also provides the key for the necessarily elaborate systems of rules essential to the prescribed protection.

Another major class of problems arises from the fact that errors and inconsistencies in the data as reported, transcribed, and coded are always discovered.

in the process of editing tabulations for publication. These errors are generally corrected at the levels of aggregation at which data are published, but often are not carried back to the basic records for the individual respondent unit, in machine sensible form. Carrying back such corrections to the basic files is prerequisite to the kind of analysis of interrelationships which is here proposed. Some method of insuring that this is done, and that the basic records meet appropriate standards as archives, is essential. Such tasks generally receive a low priority in the statistical agency whose primary task is production and publication, rather than analysis, of data.

As a technical device, the use of master samples can achieve a high degree of unification of information about the individual reporting unit. Moreover, proper experimental designs provide for the straightforward estimation of sampling variances; differences among subpopulations can be measured with specified precision if this requirement is stated in advance; variances due to differences among samples can often be eliminated; costly matching of units at the later stages can be avoided. Inconsistencies arising from many kinds of differences between surveys can also be avoided.

The use of the current population survey for special questions (veteran status, duration of unemployment, preferred number of hours, work, etc.) permits a number of useful comparisons with standard information about labor force status. Similarly, it may often be practical to use master samples to obtain information about subjects vital to some of the newer Federal welfare programs. Hence, an integrated system of master samples of households could be used to collect information about income, education, health, crime, employment, social services, housing, demography, voting registration, and the effects of opening or closing industrial plants. Not only could information be compiled about each of these subjects, but analyses could be performed which inter-related several of the subjects: education-income-crime rates, health-housing-education, etc., without loss of information or the introduction of uncertainty arising from variances between samples or from different survey practices. Indeed, the judicious use of master samples can lead economically to conformity with the general principles stated above.

The fourth principle is ability to produce any information computable from the basic records. The principle acquires new power when combined with the other three principles, because a great deal more becomes computable. The concept of "inherently computable" is taken literally, and includes kinds of statistical operations not now widely used. The paucity of current use derives from unfavorable economics, unsuitable organization of data, insufficient available detail, failure to use known techniques, obstacles growing out of confidentiality restrictions on data intended for input to the analysis, and current habits and practices deriving from all of the other obstacles. In short, the current ways of doing business fall far short of the potentiality of advanced statistical techniques applied to a well-organized body of Federal data. The present report suggests the means for mitigation or elimination of the shortcomings of the statistical system built before modern computers became available.

In these terms, the notion of "inherently computable" takes on new meaning. Obviously included are the routine computation of averages, cross-tabulations, correlations, curve fittings, time series analysis, seasonal adjustments, distribution statistics, and the application of other techniques of mathematical statistics. But it would also now be possible to test the reconciliation of one series of data against others. Test adjustments of all sorts, even very detailed and burdensome adjustments involving manipulation of very large matrices, could be countenanced. Errors could be studied, including those for whose estimation there is little theoretical foundation—the myriad kinds of inconsistencies of definition, practice, error rate, personnel, etc.—when data from two or more sources are used. Propagation of errors through the system, especially in the major synthetic series (national income and product accounts, Federal Reserve Board production indexes, price indexes) could also be studied and estimated. One set of objectives would be error detection and measurement in the primary collections. Another would be consistency testing and anomaly detection in two or more collections from different agencies, geographic regions, time periods, etc.

Computations of this sort could also be used to assist in the setting of quality standards for Federal data and for validation or certification of particular bodies of data.

There is already a praiseworthy trend toward the use of computers in the production of standard series of data by several agencies. The notion of "inherently computable" includes the generation of new series for special purposes

through the adjustment of standard series, limited only by the techniques and imagination of social scientists.

#### *Nature of the services to be rendered*

This section discusses the services which a Federal Statistical Data Center could render. The characteristics are enumerated as they are in the summary of this report. An information system capable of providing these services is described in the next section.

The services which are proposed for this suggested system are discussed below:

1. The Federal Statistical Data Center would provide data in cases where the primary agency in possession of the data is not capable of making it available in the required format, detail, flexibility, or quality. Primary agencies would continue to provide data which they can furnish in the needed form, even though they had previously delivered the relevant basic data to the Center. For example, an agency might produce statistics as a byproduct of its principal mission, having no resources to organize the information for flexible or rapid access. Or the data might require adjustment or reconciliation which the collecting agency cannot perform as well as the Center.

2. The Center would provide data where the information originates in two or more reporting systems or agencies, in order to make available information about interrelationships in maximum feasible detail, without restrictions resulting from screening for improper disclosures at the time of transfer into the Center and through association of information from multiple sources relating to the same individual reporting unit or analytical unit. The intent of this specification and its improvement over present characteristics of the Federal statistical systems are discussed in an earlier section.

3. The Center would maintain an archive of statistical data, complete in the sense described in (2) above, with all corrections and adjustments carried through in a consistent manner, and with a collection of the accompanying code-books and manuals. The intent of this item is discussed in an earlier section.

4. Outputs (responses to queries) would be provided in a variety of forms at the customer's option: printed tabulations, machine readable tapes, graphs, diagrams, etc., either locally or through telecommunications.

5. The Federal Statistical Data Center would establish, maintain currently, and operate a reference and referral service for the Federal statistical system. This service is not concerned with the actual provision of data. It deals more with those matters that a user might need before he can formulate a proper query. The reference and referral center would give information about various concepts that lie behind the statistics: general imports in contrast with imports for consumption; total employment and employees in establishment; value of product and value added; industry and product statistics, etc. Questions that could not be answered at the Center would be referred to specialists in the various agencies; the Center would identify and locate the specialists. It would also protect the experts from inquiries that could satisfactorily be managed at the Center. Personnel at the Center would be equipped with reference documents to show dates for which each kind of data is available, changes in coverage, changes in definition, changes in quality, schedules for availability of future statistics; materials available in book or report form both for data and information about their definition, method of collection, adjustment, etc. Reference services would also be provided for information not in the Federal collection: statistics from trade associations, industrial institutes, State and local governments, international organizations and foreign governments. Statistical data that can be obtained directly from the primary collection agency would be known to the Center, which would act as a referral agent for the agency.

6. ADP equipment would be available for computation and data reduction in response to queries of customers: cross tabulations, averages, distribution statistics, smoothed curves, trend fittings, seasonal adjustments, periodic analyses, correlations, regressions, and more advanced analyses in order to give access to the full range of information computable from the collection.

7. Confidential audits would be performed by machine upon the information intended for release to customers. It is recognized that this raises complex and difficult issues which require intensive study. However, there are strong reasons to believe that these issues can be resolved with the aid of modern tools of the mathematical and computer sciences.

8. ADP equipment would also support a battery of services to the statistical system of the Federal Government: computations essential to the conduct of test adjustments on statistical series and collections, computations for test reconciliations of data for two or more sources or for two or more time periods, detection of errors in primary collections or derived statistics through consistency tests and anomaly detections routines, computations necessary for the study of error propagation through the Federal statistical system, combinations of the above computations in support of validation studies for Federal statistics and in support of procedures for certification of the accuracy and consistency of Federal statistics. Much of the work referred to here is not done at present. Ordinarily, the larger synthetic statistical series are prepared by gathering data from many sources and adjusting them in various ways including their reconciliation to benchmarks of higher quality. In many cases, the source series themselves are compounded from smaller elements, sometimes in several stages before reaching down to the point of primary collection from the respondents. Computers are used for convenience and economy to speed up the processing in most of the more elaborate systems. However, in this statistical production network, there is practically no feedback of information from this process to the primary collection agencies. The adjustments required to maintain the larger synthetic series are sufficiently burdensome and closely scheduled that there is neither time nor staff for research on adjustments or the conduct of test adjustments no matter how desirable this might be in the view of the interested agencies. The combination of the comprehensive unified data system and adequate ADP equipment would create a favorable climate for this work. In addition, all manner of test comparisons across different statistical series, and many kinds of consistency tests, could readily be performed. With much of the synthesis of major statistical series on compatible computers, the effects of errors in all stages of collection, estimation, and adjustment could be studied. Hence, studies of the quality of Federal Statistics could add such techniques to existing appraisals which are based on information about the collection (completeness, sampling variance, quality checks), size of adjustments to benchmarks, and a very limited kind and number of consistency checks.

9. ADP equipment would also be used to service agencies with large-scale adjustment and reconciliation burdens (which is already being done by some agencies) in the production of standard series, and to service the creation of new series through the reconciliation and adjustment of standard series. Specialized users could define new synthetic series based upon adjustment of the standard series. However, at present, such adjustments could be applied only to highly aggregated forms of the statistics because of the cost and cumbersome nature of the process. What is contemplated here is a much more complete reprocessing designed to retain a large measure of the detail available for the standard series. For example, the input-output transactions matrices (which are now embedded in the national income and product accounts), could be transformed from the present industry-based sectoral definitions to an activity basis (in which there are no secondary products).

#### *Character and organization of the data in a Federal Data Center*

This section presents and discusses principles governing a well-integrated body of statistics arising from the work of the Federal agencies. The items discussed below are numbered as they are in the summary of this paper. While there are intimations of services that the Center might perform, there is no attempt to describe the services as such in this section; the preceding section is devoted entirely to that end. This section relates to the internal structure and operation of the Center—in matters of information—and the preceding section views the same Center from the outside, as a series of capabilities to assist the customer to obtain data.

The principles are discussed in numerical order below. While it is convenient to set forth the seven separate items for exposition and reference, the entire characterization is conceived as a single entity: no item is to be read out of its context with the other items. It is the interaction of the points taken two, three, or more at a time that characterizes this report, in contrast with possible studies of the distinct issues one at a time.

1. Principles must be developed for the initial selection and future accession of data to be included in the Center's collection; they should reflect a broad range of uses and full utilization of basic information rather than a codification of present uses, present practices, and present compromises.

The selection should recognize the importance of data acquired in the administration of regulatory programs and welfare or benefit programs. A primary purpose of the Federal Statistical Data Center is the organization of information in such way as to permit the use of data from various sources in the same analysis. The byproduct information from many agencies must now be reconsidered to determine how these data can best be combined with those of other reporting systems to contribute to socioeconomic analysis. Both program data (amount of grant, number of grantees, geographic location of program elements, etc.) and information about applicants can enrich the existing store of socioeconomic statistics from major statistical agencies.

The above paragraph is addressed to only one of the general principles discussed in an earlier section. Issues of inclusion or exclusion of various classes of data must be reviewed with all four general principles in mind and also with appreciation of the remaining items on this list itself.

2. Methodology and principles must be developed for the conversion of present files and production data to suitable archive form and quality, and their maintenance in proper form and quality, supported by codebooks, manuals, etc. It is not to be presumed that complete and consistent records will arise routinely from the collection, adjustment, analysis, and publication of data. Resources must be made available and priorities assigned. Above all, standards of form and quality must be prescribed and checked in some regular manner. Procedures must also be prescribed for work with the archive collection to prevent loss or contamination of the master records by tape erasure, statistical adjustment, aggregation or reclassification.

(a) The principles established for archives must presume that data will be transferred from the collecting agency to the Center in full available detail. The decisions about the lowest level of detail—other than the separate record for each respondent or analytical unit—will often be arbitrary. They will reflect notions of the finest detail that analytical purposes are likely to demand. In principle, there is no such ultimate disaggregation for many reporting units. For example, the use of the establishment as the reporting unit in many standard statistical systems is frequently dictated by the inability to define or obtain information for subestablishment entities; the choice is not based upon satisfaction with the level of detail obtained. The term "full available detail" must be read with a rule of reason.

(b) The principles established for archives must provide for the systematic completion of all corrections and adjustments to all data affected and all levels of detail, resulting in a fully reconciled and consistent body of data.

(c) The principles established for archives must provide for data to be transferred to the Center without screening for confidentiality; all confidentiality audits and checks would be applied to the formats and information content of the output of the Center.

(d) The principles established for archives must provide for the preservation of the identity of the reporting unit and the association of all information about the unit without regard to the agency or manner of the information; the rules and the economics of matching of existing records will be very different from those intended for future production of data.

3. Standards must be developed for definition, coding, classification, and aggregation with the intent of maximizing the ability to use different kinds of data in the same analysis, and of minimizing the loss of information. The reasons for this requirement are set forth in an earlier section of this report. A large number of standards would have to be developed, beginning with such seemingly elementary concepts as a household, a structure (a building), a business organization, an establishment (industrial), a populated place, a county; and proceeding to a school pupil, a hospital day, a recipient of (some particular) welfare service, etc. Some such standards now exist, although they are neither wholly satisfactory nor uniformly observed. These existing standards should be reexamined and many new standards developed. All standards need to be more rigorously defined and more effectively enforced. Ideally, a close matching in many dimensions of classifications, hierarchical aggregation, timing, and spatial extent should apply to all information in the basic record; but a system of practical compromises would unquestionably have to be accepted.

4. Standards must be developed for formats in order to facilitate the management, housekeeping, and retrieval of records and to avoid the loss of information. This item presumes that the characteristics of the information have, in principle, been defined. In practice, formats greatly influence the effectiveness, economy, and error rates of the whole operation.



5. Standards must be developed for quality of data (consistency of definition, error rates, etc.) and for means of assuring maintenance of quality. It is essential to know—and to issue with the statistical data—information on the quality of the data. After standards have been agreed upon, appropriate quality-control procedures would have to be instituted.

6. Automatic data processing (ADP) equipment and systems must be available to the Center to provide economy, timeliness, and flexibility of access to the information in the records. This report does not deal with configuration of computing equipment. It may be noted in passing, however, that various units might be geographically scattered. This would allow inquiries to be made from points distant from the basic record stores and the replies or outputs to be received in these same remote locations. Probably more importantly, computer laboratories in universities, research institutes, business organizations and governmental agencies could be used to transmit requests for information over long-distance lines and to receive and store information. This would provide a convenient location for trial manipulations by those making the inquiries without disturbing the rest of the communication network. Such an arrangement would give increased service and analytical power to the participating analyst.

(a) ADP must be available—

(1) to compute statistics that are inherently computable from the records,

(2) to provide answers in the required form, and

(3) to avoid unnecessary withholding of information as a result of using inefficient and redundant disclosure criteria.

Note that, for the item immediately above, the confidentiality audit would have to apply before the information was transferred to an off-line computer under the control of the user. This discussion implies that disclosure rules would retain their essentially logical character without taking into account the possibility of introducing elements of probabilistic inference in determining whether or not an undesirable disclosure might be made.

(b) ADP must be available for the conduct of confidentiality and other disclosure audits; such rules and procedures to be applied to the data in the form and content intended for release from the Center. The comment on item 6(a) applies here as well.

7. Criteria must be developed for assuring that the Center is established—and is maintained—in a manner that is responsive to a broad base of potential users, rather than in a manner which seems to suit the present habitual users at any time.

8. Criteria must be devised for periodic review of the value of the data contained in the archive followed by a selective purging of the data whose retention is no longer justified.

In sum, the main purpose of a Federal Statistical Data Center is to create a better integrated information network, for use by Government, industry, and the research community, which will provide better understanding of interdependencies within our pluralistic society, leading to better informed choices among alternative policies and programs, and more effective program implementation.

### APPENDIX 3.—THE NEW COMPUTERIZED AGE

[From Saturday Review, July 23, 1966]

Few technological developments are formidable enough to mark turning points in human history. Two such phenomena have occurred in our time: the atomic bomb and the computer.

The implications of the bomb are beginning to be understood—its capacity for instant and total destruction has been demonstrated. The implications of the computer as yet are only faintly comprehended. That they will be awesome is already apparent. Indeed, as Dr. Jerome B. Wiesner, Dean of Science at the Massachusetts Institute of Technology and former science adviser to President Kennedy, wrote recently in *The New York Times*:

"The computer, with its promise of a millionfold increase in man's capacity to handle information, will undoubtedly have the most far-reaching social consequences of any contemporary technical development. The potential for good in the computer, and the danger inherent in its misuse, exceed our ability to imagine. \* \* \* We have actually entered a new era of evolutionary history, one in which rapid change is a dominant consequence. Our only hope is to understand the forces at work and to take advantage of the knowledge we find to guide the evolutionary process."

The following special section is an attempt to identify some of these forces and to consider their implications. Nine authorities of diverse backgrounds discuss the possibilities and dangers of a computerized age. As their reports make clear, ultimately no area of human life will remain untouched by it. In the words of Automation Consultant John Diebold, whose article, "The New World Coming," introduces the section, "A complete new environment will exist."

The changes in business, government, science, education, and communications are occurring at a time when our technological capacity already has outstripped our understanding of many of its ramifications; when, as Marshall McLuhan, University of Toronto professor who often is quoted on the influence of electronic media (see Erik Barnouw's article), has said, the tumultuous pace of change already has resulted in an "information overload."

Decades ago, W. B. Yeats wrote, "The visible world is no longer a reality, and the unseen world is no longer a dream." More and more this will be true in the computerized age.

In addition to Mr. Diebold and Professor Barnouw, contributors to the section are: Gen. David Sarnoff, chairman of the board of the Radio Corp. of America; John W. Macy, Jr., Chairman of the U.S. Civil Service Commission; Patrick Suppes, director of the Institute for Mathematical Studies in the Social Sciences, Stanford University; Don D. Bushnell, associate director of the Brooks Foundation, and past president of the Association for Educational Data Systems; the Reverend Vernon F. Miller, pastor of the Goshen City, Ind., Church of the Brethren; John Tebbel, New York University journalism professor and author; and John Lear, *Saturday Review's* science editor.

The editors wish to acknowledge the valuable contribution of William L. Schubert, of the McCall Corp., in the basic planning and preparation of this issue.

—The Editors.



## THE NEW COMPUTERIZED AGE—1: THE NEW WORLD COMING

*Tomorrow's computers will revolutionize business, education, communications, science—in ways only dimly foreseen*

(By John Diebold<sup>1</sup>)

It is an extraordinary era in which we live. It is altogether new. The world has seen nothing like it before. I will not pretend, no one can pretend, to discern the end; but everybody knows that the age is remarkable for scientific research into the heavens, the earth, what is beneath the earth; and perhaps more remarkable still is the application of this scientific research to the pursuit of life. The ancients saw nothing like it. The moderns have seen nothing like it, until the present generation \* \* \* The progress of the age has almost outstripped human belief.

Those words were not spoken today—though I choose them to set today in perspective—but were used in 1847 by Daniel Webster when he opened a new stretch of railroad track in New Hampshire. A greater parallel exists between that era and our own than we normally realize. In that earlier era, science first began to be applied on a wide scale and out of that process came an entirely new society—an industrial society. Out of it, too, came problems, many of which still plague us. When we look back at that great technological upheaval, the real significance of those then-wondrous machines is the human and social change that accompanied their industrial use.

Just as yesterday's innovations proved to be moments in history—way stations leading to newer technology—so today the conception of the computer which we have learned to accept is becoming a thing of the past. Up-to-date systems are no longer glassed-in, carefully isolated accounting machines. Instead they perform an almost limitless variety of functions, and vary with individual requirements.

For example, the newest computer systems may appear as input-output units in individual desks; small televisionlike screens with keyboards and copying devices. When you ask a question you see the answer almost simultaneously on the screen. If you want a copy of the answer, you can make it immediately. The heart of the system is a switching center rather like the telephone system. Computers, storage elements of many varieties, and many other devices used as part of the system are accessible as you need them, connected through the switching center to the terminal unit at your fingertips. Thousands of people may use such systems at the same time, and each need know no more about the operation of the system than the average person knows about the telephone. In the next decade the typical computer system is going to be of this kind.

Another radical change stemming from these new computer systems involves the relationship between man and machine. One no longer need carry data down to a computer center, or go through a laborious process of getting it into the machine and then waiting for results. Each technological development is moving us toward an easier, more productive relationship between man and machine. Already, for example, a computer can transpose a rough design into exact specifications. If an engineer makes a free-hand drawing of a bridge on such a system's television-like screen, the computer will convert the drawing into exact engineering specifications, will calculate and display materials and stress, and show the design in whole, in part, or in any perspective, in immediate response to the engineer's requirements.

Looking ahead, we see important changes in technology such as chemical memories; fluid and pneumatic systems that have instantaneous response; ability to store images, graphs, drawings, and photographs, and to transmit them around the world. All these will be important elements of future computer systems. Graphic elements and the ability to communicate with TV screens are already becoming influential in progress being made in computer design. Yesterday these elements were undreamed of.

Work is being done on language translation by machine. Some document-translation is already on a regular production basis—in fact, people are now attempting to digest articles by machine. This work is still in its beginning stages and there are many problems to be overcome. But the history of this

<sup>1</sup> The author, who generally is credited with coining the term "automation," is head of the Diebold Group, Inc., management consultants.

technology is that what seems impossible today becomes an accepted part of our lives tomorrow.

Development of voice recognition by computer, while rife with problems, also is yielding results. Despite all the difficulties, voice-recognition equipment can be purchased today. No serious forecast about computer systems in the 1970's can omit voice recognition systems with several-thousand-word vocabularies. If this sounds unpromising, remember that only a few years ago people used to have 2- and 3-day meetings to discuss the problem of keeping records on magnetic tape. How naive that seems to us now. Today, we already have machines that learn (they are called heuristic machines), that devise their own route to a goal or solution; machines that recognize patterns; and machines that can devise their own strategies—for example, winning at games with the men who design them.

Adding tremendous impetus to the technological explosion is the fact that, as computer capabilities are increasing, costs are decreasing. Between 1963 and 1972—a single decade—there will be a decrease of 85 percent in the cost of completing a typical data-processing job. During this period, the cost of storage by magnetic tape will go down by 97 percent; the cost of image storage by 96 percent; and communication line costs, because of increased speeds of transmission, will decrease by 50 percent. These changes in economics will mean that we will be able to do more with information technology than we now can even imagine.

Let me turn now to the problems of putting these machines to work.

Nowhere is the turn toward technology more obvious than in the way we manage. When we first started to apply computers to business operations in 1954, we went through a very difficult experimentation period and were faced with the most puzzling kinds of problems. We have largely emerged from that period, however, and today we are using computers in business for almost everything conceivable—and much that was not just a few years ago. Senior management has begun to realize that the application of this technology is too important to leave to technicians, and that dramatic things can be accomplished if people who know the objectives of a business will take the responsibility of putting these new capabilities to work. When this happens, you find remarkable achievements.

But along with this progress have come new questions and problems. There are, for instance, union negotiation questions. Throughout the country, a number of owners of newspapers have been willing to stake the very existence of their enterprises on the right to install a computer to prepare punch tape to drive linecasting machines. Just over the horizon, it is clear that this entire process will be bypassed. Is it worth risking an enterprise on a process that is disappearing?

There are many similar questions. What kind of men, for example, should be trained as managers in the new technological environment? How do we create an atmosphere that is conducive to creative people?—for more and more of our businesses must be staffed by highly educated and creative personnel. These are only a few of the problems we face.

Most important are the human aspects. They are related to every problem we have in this field: questions of fear and uneasiness when faced with technological changes; questions of education; questions of identification with an enterprise, with a profession.

But along with the question of how we manage are questions concerning what we manage—of new areas of business opportunity. Here, I will speak of four main new entrepreneurial opportunities. The first is the obvious one that has already taken form—the industry that supplies the systems and the equipment. It is already a multibillion-dollar industry, and this is only the beginning.

The second example, as yet nonexistent but about to bloom as an important basic industry, is the data utility field. This is analogous in some ways to the electrical utility industry: It is cheaper for many people to use a central utility than for each individual to have his own generator. The same economic reasoning applies to the data utility industry, where many people can use a machine simultaneously. The technology of real-time processing, time-sharing, and communication will allow this to happen. Small- and medium-sized businesses—and for some purposes large businesses—will just plug in for data processing as we now do for electricity.

The third example is the one now being called the inquiry industry—in some ways, the publishing field of the future. This will allow the sale of proprietary data over a communications system in answer to a query placed by the customer. The possibilities are unlimited; practically any information can be provided.

We have already started to see the purchase of publishing firms by electronic companies, and this is just the beginning. There will be major changes in ownership in this area in the near future as businesses begin to position themselves to offer such services.

The fourth example is an industry of computer-based educational systems. As technology allows a dynamic or "alive" relationship between a student and a machine system that answers questions as they are posed and discerns gaps in a student's basic grasp of a subject, the much-heralded but until now disappointing teaching machines (better, I think, called learning machines) will begin to mean something. Such systems are already at work in some industrial situations—IBM's maintenance training being a good example. Other precursors can be seen in mentally handicapped children's use of computer-driven typewriters to help them overcome some of their handicaps.

If there is one salient fact about information technology, it is that it is going to produce enormous social change. As the quality of life is changed, as the rate of learning, information, travel, and communications all change, we will see a major change in living patterns, in hopes and desires. In short, a complete new environment will exist.

### THE NEW COMPUTERIZED AGE—3: NO LIFE UNTOUCHED

*By the end of the century computers will affect every field in innumerable ways; some specific predictions*

(By David Sarnoff<sup>1</sup>)

In our increasingly complex world, information is becoming the basic building block of society. However, at a time when the acquisition of new scientific information alone is approaching a rate of 250 million pages annually, the tide of knowledge is overwhelming the human capability for dealing with it. So man must turn to a machine if he hopes to contain the tide and channel it to beneficial ends.

The electronic computer, handling millions of facts with the swiftness of light, has given contemporary meaning to Aristotle's vision of the liberating possibilities of machines: "When looms weave by themselves, man's slavery will end." By transforming the way in which he gathers, stores, retrieves, and uses information, this versatile instrument is helping man to overcome his mental and physical limitations. It is vastly widening his intellectual horizon, enabling him better to comprehend his universe, and providing the means to master that portion of it lying within his reach.

Although we are barely in the second decade of electronic data processing, the outlines of its influence on our culture are beginning to emerge. Far from de-personalizing the individual and dehumanizing his society, the computer promises a degree of personalized service never before available to mankind.

By the end of the century, for the equivalent of a few dollars a month, the individual will have a vast complex of computer services at his command. Information utilities will make computing power available, like electricity, to thousands of users simultaneously. The computer in the home will be joined to a national and global computer system that provides services ranging from banking and travel facilities to library research and medical care. High-speed communications devices, linked to satellites in space, will transmit data to and from virtually any point on earth with the ease of a dial system. Students, businessmen, scientists, government officials, and housewives will converse with computers as readily as they now talk by telephone.

In the health field, computers will be employed to maintain a complete medical profile on every person in the country from the hour of birth. The record will be constantly updated by a regional computer for immediate access by doctors or hospital personnel. The computer also will maintain files on every known ailment, its symptoms, diagnosis, and treatment. A doctor will communicate a patient's symptoms to the computer center and within seconds receive suggestions for treatment based both on the symptoms and the patient's history.

Computers will handle the Nation's fiscal transactions from a central credit information exchange, to which all banks, business enterprises, and individuals will be connected. Purchases will be made, funds invested, and loans issued by transfers of credit within the computer without a dollar or penny physically

<sup>1</sup> Gen. David Sarnoff, who this year is celebrating his 75th birthday anniversary, is chairman of the board of the Radio Corp. of America.

exchanging hands. Even the soil will be computerized. The long-range outlook for agriculture includes new sensing devices that will be placed on larger farms, feeding information to the computer on soil moisture, temperature, weather outlook, and other details. The computer will calculate the best crops to plant, the best seeding times, the amount of fertilizer, and even the correct harvesting time for maximum yield.

Some of the most profound changes wrought by the computer will be in education. Here, the machine will do more than assist students to solve problems and to locate up-to-date information: It will fundamentally improve and enrich the entire learning process. The student's educational experience will be analyzed by the computer from the primary grades through university. Computer-based teaching machines, programmed and operated by teachers thoroughly trained in electronic data processing techniques, will instruct students at the rate best suited to each individual. The concept of mass education will give way to the concept of personal tutoring, with the teacher and the computer working as a team. Computers will bring many new learning dimensions to the classroom. For example, they will simulate nuclear reactors and other complex, dangerous, or remote systems, enabling students to learn through a form of experience what could formerly be taught only in theory.

The computer's participation in the field of learning will continue long after the end of formal education. The government estimates that 50 percent of the jobs to be held 10 years from now do not even exist today. With this tremendous rate of occupational obsolescence, future generations of Americans may pursue two or three careers during their lifetimes. The home computer will aid in developing career mobility by providing continuing self-instruction.

Just as it is recasting the educational process, the computer is also fundamentally changing the production and distribution of the printed word. Five centuries ago, Gutenberg broke words into individual letters. Electronic composition now breaks the letters into tiny patterns of dots that are stored in the computer's memory. Any character can be called up by the computer, written on the face of a cathode ray tube, and reproduced on film or paper in thousandths of a second. Nothing moves except the electrons.

When the electronic computer first appeared in composition rooms and printing shops several years ago, its job was to hyphenate words and justify text. But the computer, working at speeds of thousands of words a minute, was driving mechanical typesetting devices capable of setting only a few words per minute. Now, the development of computerized composition makes it possible to set text at hundreds of lines per minute. Photographs and drawings will be set the same way. Since the printed picture is itself a dot structure, the computer can electronically scan any photograph or drawing, reduce it to dots and store it, then retrieve it and beam it on a cathode ray tube for immediate reproduction.

In the future, electronics will develop processes that will make it possible to go from final copy and illustrations to printing in one integrated electronic process. One result will be that newspapers, in the foreseeable future, will no longer be printed in a single location. Instead, they will be transmitted through computers in complete page form to regional electronic printing centers that will turn out special editions for the areas they govern. Local news and advertising will be inserted on the spot. Eventually, the newspaper can be reproduced in the home through a small copying device functioning as part of a home communications center.

Basic changes also will come to other areas of the printed word. For example, of the more than one billion books published every year, almost half are textbooks. The growth of knowledge and the factor of obsolescence mean that these texts must be supplemented by a professor's mimeographed notes. Today, these notes have a small distribution of only a few hundred copies. Computers will make it possible to catalog this information and thus broaden its availability.

At the turn of the century, most large universities will not only have electronic composition systems that allow them to reprint original research, theses, or course notes upon demand; they will also have a computerized information retrieval library. This process of information retrieval can be duplicated in almost any other field. The scientist will have the latest technical papers culled by the computer and reproduced in the laboratory or home. The computer will bring to the attorney all the pertinent laws, decisions, and precedents on any case that concerns him. The business executive need not rush to the office every morning; most of the information he will need to conduct his business will be run off for him at home, and he will have a two-way national and global closed-circuit television, via satellites, for meetings and conferences.

Some of these developments are probabilities, some of them are certainties, and all of them are or soon will be within the capabilities of the computer art. But one fact is absolute: the incredible growth of the computer in numbers, power, and availability.

In just 10 years, the typical electronic data processor has become 10 times smaller, 100 times faster, and 1,000 times less expensive to operate. These trends will continue, and our national computing power, which is doubling every year, will soon be sufficient to make the computer a genuinely universal tool.

In 1956, there were fewer than 1,000 computers in the United States. Today, there are 30,000, or more than \$11 billion worth; and by 1976 the machine population may reach 100,000. And these figures will, of course, be greatly increased through the growth of data processing in other nations.

A decade ago, our machines were capable of 12 billion computations per hour; today, they can do more than 20 trillion, and by 1976—a decade from now—they will attain 400 trillion—or about 2 billion computations per hour for every man, woman, and child. Quite evidently, the threshold of the computer age has barely been crossed.

Nevertheless, for all its potential to stretch the mind a thousandfold, it is perhaps necessary to point out that the computer is still a thing—that it cannot see, feel, or act unless first acted upon. Its value depends upon man's ability to use it with purpose and intelligence. If his postulates are wrong, the computerized future can only be a massive enlargement of human error.

Ramsay MacDonald once warned against "an attempt to clothe unreality in the garb of mathematical reality." Computers echo this warning. For they cannot usurp man's unique ability to blend intuition with fact, to feel as well as to think. In the end, this remains the basis of human progress.

The task ahead will be to assign to the machine those things which it can best do, and reserve for man those things which he must provide and control. It is my conviction that society will adjust itself to the computer and work in harmony with it for the genuine betterment of life.

#### THE NEW COMPUTIZED AGE—9: WHITHER PERSONAL PRIVACY?

*Computer Technology May Enlarge Man's Liberty or Inhibit It; New Rules Must Be Made; New Questions Answered.*

(By John Lear)

Between my resignation as an editor of the slowly dying Collier's and the inauguration of SR's Science and Humanity Supplement, I was for a short while a minor adviser to Thomas J. Watson, Jr., son of the founder of the International Business Machines Corp. My experience there contradicted two popular beliefs about IBM.

The first belief was that all IBM employees were required, as a condition of employment, to wear white shirts on the job, stay sober at home, and maintain upright on their desktops identical copies of a small sign bearing the personal command of Thomas J. Watson, Sr.: "Think." I wore light blue shirts, drank cocktails at lunch, and put the "Think" sign on the windowsill of my office whenever my secretary put it back on my desk; yet I had a standing invitation to young Tom's sanctum.

The second of my working conditions that ran against supposed IBM tradition had to do with that then-new phenomenon, the so-called "giant brain," or high-speed electronic computer. IBM has sold an enormous number of electronic computers. According to legend, a loyal IBM salesman would leave his wife if necessary to clinch another sale. The truth about IBM computer sales, as I experienced it, was that an immense share of IBM's sales investment went into persuading eager customers to delay the purchase of computers.

Delaying was sound business practice because a computer can do only what it is told to do; it must follow instructions literally; until the instructor himself is sure where literal pursuit of a long series of tiny steps will lead, turning the task over to a computer can be dangerous. The machine may complete its assignment before its owners realize that the outcome isn't really the one they seek.

The period of my IBM experience dates back roughly a dozen years. At that time, computers could do only one thing at a time, in sequence. Computer programmers—the people who break everyday English into binary arithmetic messages (consisting entirely of numerical zeros and ones) comprehensible to the ma-

chines—were such valuable individuals that IBM allowed them to sit at the computer control consoles as long as necessary to figure out flaws in the translation process. Since then, the speed of the machines has risen, the size of components has shrunk, and computer sophistication has grown to such a state that a machine can do a number of different things at a time, not necessarily in sequence.

As a result, the computer's time is worth upward of \$300 an hour—too valuable to be wasted by a programmer's headscratching; and each programmer now must figure out unexpected problems in a separate place while the computer goes on with other people's problems. In other words, it is now possible for wrongly instructed computers to make more disastrous and far-reaching mistakes in a shorter time than ever.

Paul Baran, of the Rand Corp. in Santa Monica, has studied this dilemma more searchingly than most observers of the phenomenon.

"As we pass through life," he reminds us, "we leave a trail of records, widely dispersed and generally inaccessible—except with a great deal of effort and diligence. Beginning with a birth certificate, we accumulate hospital and medical records. We become deductions on our parents' income tax. In school, we generate records of our grades, attendance, IQ tests, personality profiles, etc. (Automated teaching will add to this recordkeeping. The volume of data recorded per child may be expected to increase even more markedly.) After school we start accumulating employment, social security, and selective service records. We may get a driver's license. Most of us will apply for marriage licenses, and some of us will collect divorce decrees which will end in voluminous court records. If we are lucky, we will be able to avoid having arrest and jail records.

We move from job to job in a mobile economy creating moving-company inventory records of our goods. Even as we move from place to place we leave behind short records of our airplane reservations and, for some reason, every hotel makes a ritual of acquiring and preserving the alleged names and addresses of its guests for posterity. This is only a partial list. Think of all the records you leave as you go through life.

"Behind all this creating of records is the implicit assumption that they will some day be of use. In order to be of use, there must be some means of interrogating the files to resurrect the information sought.

"An Internal Revenue Department investigator might wish to have immediate access to the tax returns of each of the associates of a man who is being audited, in order to check on consistency of financial relationships.

"A company may wish to have rapid access to its personnel files to know whether to give a good reference to a former employee.

"A doctor may wish to trace the entire medical history of a patient to provide better input into a diagnostic computer.

"The Veterans' Administration may wish to examine a man's complete military record and possible other previous medical records to see whether the ailment claimed as being service connected really is service connected.

"A lawyer for the defense of a man will wish to search for jail and arrest records, and possibly credit records of all witnesses for the plaintiff.

"Professional licensing boards may want to delve into any records to determine if an applicant has an unblemished character.

"The military in filling extremely sensitive positions may even wish a record of all books borrowed by a prospective applicant to insure that his interests are wholesome and he possesses the proper political bias desired.

"Today it is difficult to gather such information about a prospective examinee. If one went through direct channels and asked most sources for their records about a person, he would most likely be rejected, if for no other reason than that the information is not available—cheaply. Even if the records were publicly available, the investigator would have to spend a great deal of time and effort delving through to discover pertinent data. Today, as a practical matter, if one wishes to obtain certain information about a person, he hires a private detective who charges a great deal of money and expends a great amount of time obtaining a little information available from a portion of these potential records. The price for a fishing expedition for information is high and most of the fish are inaccessible."

Having thus summed up the "the pleasant past," Rand Analyst Baran looks into the future through a three-step review of established processes of computer storage of information. Step 1: Manual records are kept by human clerks. Step 2: Some of the clerks are eliminated by putting all the records into a central

computer file with readout of the records controlled from a single point. Step 3: Information is read into and out of the file from a large number of different points.

Baran envisages connection of one remote-access computer with other similar computers, and through this, "danger of loss of the individual's right to privacy as we know privacy today." "The composite information data base may be so large and so easily accessible that it would permit unscrupulous individuals to use this information for unlawfull means," he warns. "Modern organized crime should be expected to have the financial resources and skills necessary to acquire and misuse the information."

He expresses concern not only over the possible creation of "automated black-mail machines" but over the potential addition of "inferential relational retrieval techniques" now being developed which, "when fully refined, could determine relationships of any person, organization, event, etc., to any other person, organization, or event." Noting that "humans, by their day-to-day necessity of making decisions on totally inadequate evidence, are innately prone to jump to conclusions when presented with very thin chains of inferred relationships," he predicts an increase in the already growing practice of unearthing defaming information about candidates for political office.

The Baran forecast of computer hazards is fortified by the studies of another Rand researcher, M. R. Maron.

"Consider," suggests Maron, "what could happen as machines are used to make decisions about people. For example, consider a situation where a computer is programed to decide who should get a security clearance from the Government, or who should get an education loan, or whether someone's driver's license should be suspended, or who should get a passport, or who should be accepted for the Peace Corps or the Job Corps, etc."

"As larger files (of machine-language data, stored in computer memories, linked cross-country by telephone) become accessible there will be a natural tendency to use machines for the automatic selection (or rejection) of people according to some preprogramed set of criteria. Supposedly these criteria will have been carefully thought out before programing the machine. Even so, the implications are dangerous."

"In such a mechanized situation, how does an individual get an opportunity to 'tell the system' that its selective criteria don't apply to his own special case? Each individual is different, each has certain extenuating circumstances, each has information which he believes to be relevant to the selection decision and which the system does not consider relevant. And so on. If an individual does not have the opportunity to be judged on the circumstances of his own special (individual) situation, then he is being treated as a machine."

"Will there be a tendency in the future to create an environment where we treat each other as machines; i.e., where there is no opportunity to 'change the system's mind'? How can we create a society where we treat our citizens as people and not as machines? How can we create a society where each individual has the opportunity to explore and unfold his own special potentials—to realize what he is?"

"These questions lead to further questions—to questions about who we are and what it means to be a person. And this brings us to the problem of values. What kind of a life do we want? What kind would we value—ought we to have? How can we create a society that fosters those actions and goals that we value? How define and explicate values? How measure and compare and rate values? How select among competing values? How can we estimate the impact of computers on our values?"

"And if our projections into the future suggest that we are heading toward a future society which is not conducive to a 'good' life, what can be done to isolate the trouble spots and to influence those changes that will prevent the possible 'evils'? Such analysis of future prospects implies prediction, evaluation, and then some attempt at control. Can the process of control be made democratic so that a small professional elite does not dominate in influencing the shape of the future?"

"Finally, there is the problem of time—the time that it takes to initiate and complete corrective action. Given an analysis of the impact of computers on society and given some corrective action that must be taken in order to avoid some future situation, how long a timelag will occur between corrective action and modification of the situation?"

The positive cultural potential of computers was emphasized last January in a report to President Lyndon B. Johnson by the National Commission on Tech-

nology, Automation, and Economic Progress. Although this report dissented from the "almost" \* \* \* commonplace (opinion) that the world is experiencing a scientific and technological revolution" of sufficient power "to make our economic institutions and the notion of gainful employment obsolete," it proposed serious consideration of development of a computerized "system of social accounts" capable of analyzing accurately in advance the benefits and costs of any sociopolitical experiment. Such a system theoretically could grapple competently with complex problems such as water and air pollution, urban blight, the transportation tangle, integration of the Negro into American society, and the continuing spread of crime.

The Presidential Commission report defined the phrase, "system of social accounts," to include mixtures of systems analysis, simulation, and operations research in proportions required for particular cases. Systems analysis and operations research are now in wide employment in military planning and extraterrestrial space exploration. Simulation techniques are part of current plans for global weather observation and forecasting.

How close are we to a workable "system of social accounts"?

One of the best informed men on earth on the subject of computer development is Dr. Cuthbert Hurd, chairman of the board of Computer Usage Co., Inc. In addressing the National Automation Conference of the American Bankers Association in Chicago last month, Dr. Hurd observed that no computed manufacturer today markets an "operating system" flexible enough to apply all the diverse talents of computing machines to any complex problem.

"I suppose," Dr. Hurd told the bankers, "that as much as 200 man-years of effort might be required to produce a modern operating system, costing say \$5 million."

If such a system were to be perfected, Dr. Hurd said, "it is still unclear whether proprietorship (of the system) could be maintained under the existing patent or copyright laws."

There are two ways, then, to state the challenge of computerized society. One was succinctly put in a recent issue of the American Scholar by Lynn White, Jr., professor of history at the University of California in Los Angeles: "Must the miracle of the person succumb to the order of the computer?" The other statement comes from Paul Baran: "What a wonderful opportunity awaits us to become involved in such problems as to exercise a new social responsibility."



APPENDIX 4.—SPEECH BY VICE ADM. H. G. RICKOVER, U.S. NAVY,  
ENTITLED "LIBERTY, SCIENCE AND LAW"

[Copyright 1966, H. G. Rickover]

LIBERTY, SCIENCE, AND LAW, BY VICE ADM. H. G. RICKOVER, U.S. NAVY, AT THE  
ATHENS MEETING OF THE ROYAL NATIONAL FOUNDATION ATHENS, GREECE, JUNE  
2, 1966

This speech reflects the views of the author and does not necessarily reflect the views of the Secretary of the Navy or the Department of the Navy.

I deeply appreciate your invitation to address this meeting. It is an honor and a moving experience—especially for an American—to speak here where the ancient *Ecclesia* had its seat, where men first practiced the difficult art of self-government, succeeding brilliantly for a time but failing in the end. My country, as you know, picked up the torch of liberty they had lighted and established the first representative democracy in modern times, even as Athens had established the first direct democracy in all history.

Twenty-four centuries separate these two great innovative acts in time, over 5,000 miles in space. One took place in a small city-state possessing few material resources, the other in a huge country of great natural wealth. Yet there is a close inner link between them. They had the same objective. The principles they adopted to achieve their purpose were similar. Both sought to create—and did create—the political framework for a society of free men.

Even as Solon, Cleisthenes and Pericles before them, the framers of the American Constitution of 1789 were political thinkers, as well as experienced practical politicians. They drew upon Greek political theory and practice with which they were thoroughly familiar, adopting what had proved successful, ingeniously improving where the earlier structure had shown weakness. They were men of the enlightenment, when classical rationalism sparked a new Age of Reason throughout the Western World; when philosophers were inspired to mount an attack on every custom and institution that shackles the mind of man and arbitrarily restrains his actions—from superstition to class privilege, from tyranny by an established church to tyranny by a secular autocrat. The political institutions of all the nations of the free world today—beginning with my own—had their inception in the turmoil of that last phase of the Renaissance.

Western civilization is set apart from civilizations elsewhere, both past and present, by its dynamism, its extraordinary creativity, its intense preoccupation with things of the mind. All this started with the Renaissance. Not until modern Western man rediscovered and retrieved his classical heritage did he begin to outstrip the rest of the world.

To borrow a Churchillian phrase, it can be said of Athens, of Greece in general, that never before or since did so few human beings leave so deep and lasting an imprint on so many others, differing in race and faith, distant in time and space from this cradle of Western civilization. Their mark is on all our science, our art, architecture, literature, theater, and on our political thinking and practice as well. Here in this city, on this hill where I am privileged to stand, the Athenians proved that free men could govern themselves; that it was possible to live in a civilized society without having to relinquish personal freedom.

This was an epochal achievement. In all his long life on earth, man has had but brief moments of freedom. His own nature is the cause of the paradoxical situation that *civilization and liberty are interdependent, yet at the same time antithetical*. One cannot be had without the other, yet reconciling them remains to this day what it has always been—the most difficult political, social and economic problem.

Civilization and liberty are *interdependent* because basic to freedom is exercise of mind and spirit, of the faculties that set us apart from other living things and make us fully human. For this there must be a modicum of leisure which

comes only with civilization, when men no longer need devote all their time and energy to appeasement of hunger and protection against the elements—as must the animals.

But release from endless toil for mere survival does not automatically set men free. Indeed, the very opportunity to cultivate mind that spirit which civilization opens to man lies at the root of the *antithesis* between civilization and liberty, for this opportunity is not seized to the same degree by everyone.

Always and everywhere, civilization results in much greater enlargement of the scope of human thought and action among the minority possessing high intelligence than among the majority of average people. Nature endows men with unequal capacities for acquiring knowledge and competence. More so in the realm of the intellect, which is all-important in civilized life, than in the realm of physical strength and courage, which counts most in primitive society. Men become, as it were, more unequal as civilization advances.

When life is simple, it can be understood by nearly everyone, and the competencies needed to function effectively are within the grasp of all. This makes for the rough equality of status that is so favorable to mutual respect of one another's personal liberties. There can be no freedom unless it is mutually conceded.

With civilization, life grows complex, harder to understand for ordinary people, demanding competencies many are unable to acquire. In understanding and competence, the gifted swiftly forge ahead. What they achieve is beyond the capacity of the average. The result is that men grow apart, their interests diverge. Society tends to divide into segments according to superiority of competence or superiority of numbers. The temptation is great for each segment to use the power its particular superiority confers to bend the whole of society to its will, thus putting an end to freedom.

The Athenians were first to devise a political system that preserved the citizen's liberty by counteracting the natural human inequalities which are the root cause of segmented power centers. So precise and clear was their thinking, that the basic principles of their system remain to this day the best protection of individual freedom. Government of the people, by the people and for the people was their great invention; political equality their crowning achievement. They inaugurated the reign of mind over force by providing for resolution of differences in point of view and interest through public dialog leading to consensus, instead of by the exercise of power. Perhaps the most remarkable feature of their policy was that it engaged the continuous participation in public business of a large part of the citizenry—somewhere between one-fifth and one-fourth at any given time. It was obvious to them that only when the people are personally involved with their government will public officials be responsive to the popular will. Citizens who shunned public service were called *idiotes* and considered useless; in some cases, failures to do one's public duty resulted in loss of civil rights.

To quote Edith Hamilton, the American classicist who was made an honorary citizen of Athens, "the idea of the Athenian state was a union of individuals free to develop their own powers and live their own way, obedient only to the laws they passed themselves and could criticize and change at will." This is the political ideal that to western man spells liberty and that is rejected in toto by all autocracies, modern as well as ancient.

The Greeks, I think, understood better than most of us what it means to be free. In his play, "The Persians," Aeschylus who fought at Marathon puts his finger unerringly on what distinguished free Greece from unfree Persia. He has the Queen of Persia ask about the Athenians: "Who is their master?" To which she received the answer, "they are not subject to any man"; they obey only the law. When she is told of her son's defeat, the Queen remarks: "Even if he fail, there is no law can call him to account." How better could one express the contrast between the protagonists in today's cold war?

Action wrote that "power corrupts and absolute power corrupts absolutely." The Greeks penetrated more deeply and saw that power erodes man's reason. One senses cool contempt in Herodotus' report of the wrath of Xerxes when the bridge he ordered built across the Hellespont was torn apart in a storm. Straightway he "gave orders that the Hellespont should receive 300 lashes, and that a pair of fetters should be cast into it," and he "commanded that the overseers of the work should lose their heads." Here stands revealed the totalitarian mind—the same today as in the past.

Liberty, never gained without enormous effort and sacrifice, is all too easily lost. Those who enslave their own people seem irresistibly driven to extinguish



freedom everywhere. When we understand them, we are better prepared to ward off their aggression. More important still is awareness of the forces *within* free societies that endanger liberty. In both respects, there is still much we can learn from the Athenians.

It seemed to me, therefore, that the setting here would be eminently suited to a discussion of certain developments in modern democracies that have an adverse effect upon the liberties of the individual and the social and moral values cherished by free men. The causative factor of this new threat to liberty is science and science-based technology.

This new science-technological threat is but the latest version of the age-old conflict between civilization and liberty—a conflict that has no permanent solution but reappears perennially in new form.

Liberty is never gained for once and for all. Each generation must win it anew. Each must defend it against new perils. These perils arise because men, being endowed with free will, continually alter the conditions of life. Countless decisions made in pursuit of private objectives may so transform society that institutional safeguards once adequately protecting human liberty become ineffective. It is then necessary to return to first principles and to adapt them to altered circumstances.

The title of my speech "Liberty, Science and Law" expresses my conviction that unless certain practices in the technological exploitation of scientific knowledge are restrained by law, they will cost us our liberties.

Science and technology are, of course, of immense benefit to man. They are so highly regarded that no one would, or for that matter could, prevent their spreading to areas that at present are retarded in this respect. But they may bring about changes in our physical environment of greatest potential danger. Certain technologies admittedly injure man, society, and nature. Yet, even in countries where the people are sovereign and where they recognize the danger, efforts to bring these technologies under social control have had little success. Those who have the use of technology are powerful enough to prevent legal restraint, the main prop of their power being the esoteric character of modern science.

Much of it is incomprehensible even to intelligent and educated laymen. When scientific-technological considerations enter into public issues—as is often the case today—the issues cannot be understood by the electorate, frequently not even by the public officials who are directly concerned. There is then no recourse but to call on scientists for expert advice. In effect, the issue will be decided by them, yet they have not been elected, nor are they accountable to the people. What is left of self-government when public policy no longer reflects public consensus? And, when the public finds that it cannot judge and evaluate issues involving science, will it not become apathetic toward all public issues? Does this not spell the doom of self-government, hence of freedom for modern man? Though all the institutions established to safeguard his liberties may remain intact, the substance of freedom will have been lost.

By one of those ironies of fate beloved of Greek dramatists, this new threat to liberty has its source in the noblest Greek achievement, the freeing of the human mind to roam at will in pursuit of truth and knowledge. All things are to be examined and called into question, said the Greeks. Unless men understood the world in which they lived, and because of this understanding felt at home in it and could be useful citizens, they were not truly free. Never before or since was intellectual freedom valued so greatly. "All things were in chaos when mind arose and made order," said Anaxagoras, the mathematician and astronomer.

Everywhere else, the domain of the intellect was the special preserve of powerful priesthoods who jealously guarded their monopoly of knowledge. "To teach the people so that they would begin to think for themselves would destroy the surest prop of their power," wrote Edith Hamilton. "Ignorance was the foundation upon which the priest power rested." The legends of most people are replete with stories of divine punishment for trying to know more than was deemed proper—clear evidence of the determination of this priestly elite to discourage ordinary people from seeking knowledge. Not so in Greece. There curiosity and search for knowledge were held to please the gods, for through these the marvels of the gods were revealed to man. Wisdom and intelligence had their own protective deity—Athena.

When Renaissance man recovered his classical heritage, the most precious treasure he found was freedom of the mind. With his mental powers set free, it took him but three and a half centuries to build on foundations laid in classical

foreseen that in its ultimate consequences the scientific revolution might digreece the whole magnificent edifice of modern science. No one could have minish human liberty.

But it has brought us back full cycle. Science—the vital area of knowledge today—is for most of us virtually a closed book; again it has become the monopoly of a small elite. This is not the fault of the scientists. Unlike ancient priesthoods, they have no wish to bar others from knowledge or to use it to enslave the ignorant. Many scientists make strenuous efforts to explain science to the lay public. Nevertheless, we find ourselves in much the same position as the ancient Egyptians whose very lives depended on knowing when the waters of the Nile would rise and fall—knowledge possessed by their priesthood alone.

As in the past, it is not the knowledge gap per se that is most detrimental to freedom, not the fact that the majority cannot follow scholars into the realm of higher mathematics and science; rather it is the effect ignorance of science has on public attitudes toward science and science-based technology. The impact of technology, in particular, on the individual and on society at large is profoundly affected by prevailing concepts of what technology is and what purpose it should serve.

If people understood that technology is the creation of man, therefore subject to human control, they would demand that it be used to produce maximum benefit and do minimum harm to individuals and to the values that make for civilized living. Unfortunately, there is a tendency in contemporary thinking to ascribe to technology a momentum of its own, placing it beyond human direction or restraint—a tendency more pronounced in some countries but observable wherever there is rapid technological progress.

It manifests itself in such absurd statements as that technology *demand*s some action the speaker favors, or that "*you can't stop progress.*" Personalizing abstractions is a favorite means of semantic misdirection; it gives an air of authority to dubious statements. Most people are easily pressured by purveyors of technology into permitting so-called progress to alter their lives, without attempting to control it—as if they had to *submit meekly to whatever is technically feasible*. If they reflected, they would discover that not everything hailed as progress contributes to happiness; that the new is not always better, nor the old always outdated.

The notion is also widespread—doubtless fostered by users of technology—that, having wrought vast changes in the material conditions of life, technology perforce renders obsolete traditional concepts of ethics and morals, as well as accustomed ways of arranging political and social relationships. Earnest debates are currently taking place whether it is *possible* to act morally in the new technological society, and proposals have been made—quite seriously—that science must now *replace* traditional ethics! We have here a confusion that must be cleared up.

Through technology we are relieved of much brutal, exhausting, physical labor, as well as boring routine work; we are provided with numerous mechanical servants who do certain kinds of work faster, cheaper, and more efficiently than people. Why should the ease and affluence technology makes possible affect moral precepts that have guided Western man for ages? This may brand me as old fashioned but I have not yet found occasion to discard a single principle that was accepted in the America of my youth.

Technology is tools, techniques, procedures, things; the artifacts fashioned by modern industrial man to increase his powers of mind and body. Marvelous as they are, let us not be overawed by these artifacts. Certainly they do not dictate how we should use them nor, by their mere existence, do they authorize actions that were not anteriorly lawful. We alone bear responsibility for our technology. In this, as in all our actions, we are bound by the principles governing human behavior in our society.

Does it make sense to abandon principles one has lived by because he has acquired better tools? Tools are for utilizing the *external* resources at our disposal; principles are for marshaling our *inner*, our human resources. Tools enable us to alter our physical environment; principles serve to order our personal life and our relations with others. The two have nothing to do with each other.

This should be obvious, but erroneous concepts of science and technology abound because people tend to confuse the two. Not only in popular thinking but even among the well-informed, science and technology are not always clearly distinguished. Characteristics pertaining to science are frequently attributed to technology, even as science itself is confounded with ethics.

*Science* has to do with discovering the true facts and relationships of observable phenomena in nature, and with establishing theories that serve to organize masses of verified data concerning these facts and relationships. By boring into the secrets of nature, scientists discover keys that unlock powerful forces which can be made to serve man. It is through *technology* that these forces are then put to human use.

*Science is a body of systematized knowledge; technology is the apparatus through which knowledge is put to practical use.* The difference is important.

Because of the care scientists take to verify the facts supporting their theories, and their readiness to alter theories when new facts prove them imperfect, science has acquired great authority. What the scientific community accepts as proven is not questioned by the public. No one disputes that the earth circles the sun, or that atomic fission produces energy.

Technology cannot claim the authority of science and is therefore properly a subject of debate, not alone by experts but by the public as well. Little thought is customarily given to the possibility of harmful aftereffects by those responsible for technological exploitation of scientific knowledge. In consequence, technology has proved anything but infallibly beneficial. Indeed, much damage has been done because no thought was given to the interaction of technology with nature. More of this presently.

A certain ruthlessness is encouraged, in the mistaken belief that to disregard human considerations is as *necessary* in technology as it is in science. The analogy is false.

Rigorous exclusion of the human factor is *required* by the methods of science. These were developed to serve the needs of scientists, whose sole interest is to comprehend the universe; to know the truth; to know it accurately and with certainty. The searcher for truth cannot pay attention to his own or other people's likes and dislikes, or to popular ideas of the fitness of things. What he discovers may shock or anger people—as did Darwin's theory of evolution. But even an unpleasant truth is worth having; besides one can choose not to believe it. *Science, being pure thought, harms no one.*

Technology, on the other hand, is *action*, often potentially dangerous action. Never has man possessed such enormous power to injure his fellow humans and his society as has been put into his hands by modern technology. This is why technology can have no *legitimate* purpose but to serve man—man in general, not merely some men; future generations, not merely those who currently wish to gain advantage for themselves; man in the totality of his humanity, encompassing *all* his manifold interests and needs, not merely some one particular concern. Technology is not an end in itself; it is a means to an end, the end being determined by man himself in *accordance with the laws prevailing in his society.*

A word may be in order concerning the disparate meaning of the term *law*, depending on whether it is used in the ordinary sense—which is also the original sense of the word—or by scientists.

Law, as commonly understood, refers to those rules of human conduct prescribed and enforced by society. Its purpose is to resolve human conflicts by the application of definitive rules. These rules are always debatable and can be changed when there is demand for a change.

The scientists have appropriated the term law to describe regularities exhibited by physical phenomena—the rules by which the universe governs itself. In the transition, the word has taken on a new meaning.

From the layman's point of view, what the scientist calls law is fact, rather than law—immutable fact. Or, if you prefer, it is law operating in a sphere where human beings can exercise no influence. We cannot alter the laws of the cosmos; we can only discover them. A law of science expresses mechanical regularity where no choice of action, no free will comes into play; it deals with constancy of behavior in nature. It has relevance for us because it makes the universe comprehensible and so enables us to utilize the forces of nature for human purposes.

We are bound by the laws that science has disclosed when we exploit these forces by means of technology. Likewise we are bound by the manmade laws of our society, for our actions affect fellow human beings. Technology straddles, as it were, the law of the universe and the law of man; it is subject to both.

Much confusion in popular thinking arises from this fact. The two laws are confounded. Or, to put it differently, *they are thought to be part of a single system of law so that one or the other must perforce take precedence.*

Ever since scientists discovered that the earth is not the center of the universe, as had been maintained by the highest human authorities, we have been learn-

ing painfully that the laws of nature cannot be overturned by human fiat. It has taken a long time to attain this rational attitude; we are now conscious of the consequences of intolerance in the past. Perhaps this is why we are so tolerant toward those who claim the right to use technology as they see fit, and who treat every attempt by society to regulate such use in the public interest as if it were a modern repetition of the persecution of Galileo.

The right to be protected by law against injurious action by others is basic to civilized society. Yet, opponents of legislation intended to restrain use of potentially dangerous technologies are often able to prevent or delay enactment of such laws by playing upon the layman's respect for science. It is their common practice to argue as if at issue were a law of science when, in fact, what is being considered is not science but the advisability or legality of the technological exploitation of a scientific discovery. The public would not be deceived by such arguments if it clearly understood the fundamental difference between science, which is *knowledge*, and technology, which is *action* based on knowledge.

To guard against being misled, one should cultivate an attitude of skepticism whenever the word "*science*" is used. Is it science that is being discussed or is it technology? If technology, the question at once arises whether the proposed action is legally permissible and socially desirable. These are matters that lie outside the domain of science. Just as the law of the cosmos cannot be overturned by human fiat, so is human law supreme within its own proper sphere of operation. Technology must therefore conform to that most basic of all human laws, the maxim of the "mutuality of liberty," the principle that one man's liberty of action ends where it would injure another. Without this maxim, freedom would be a barren privilege.

Whether or not a particular technology has harmful potentialities should not be decided unilaterally by those who use it. For the user, destructive technologies are often highly profitable. He is, therefore, an interested party to the conflict between private and public interest that every potentially harmful technology poses. Nearly always he is also a practical man.

I think one can fairly say that the *practical* man's approach to a new scientific discovery and its technological exploitation is *short-range* and *private*, concerned with ways to put scientific discoveries to use in the most economic and efficient manner. Rarely will he give thought to the *long-range* and *public* consequences of his actions, that is, to the effects that a new technology may have on people, on the nation, on the world; on present and future generations.

To illustrate the disastrous consequences of a narrow practical approach, let me give some examples of technological damage to our national environment.

Carelessly emitted, the waste products of new technologies create a massive problem of soil, water, and air pollution. We may be permanently damaging the atmosphere by changing its chemical composition. New products, profitable to manufacturers and useful to consumers, are often themselves intractable pollutants. For instance, detergents which unlike soap do not dissolve in water, or pesticides and weed killers which, carelessly applied, will poison soil, crops, birds, animals, fish, and eventually man.

Other technologies enable man to alter the very contours of the land—as with new strip mining machinery. Because it cuts the cost of extraction, such machinery is used in some places. Huge chunks of earth and rock with their topsoil and vegetation are gouged out, changing fertile country into a desolate lunar landscape—a land robbed not only of its irreplaceable mineral wealth but of its fertility as well.

Man now has the means to slaughter all the wild animals on earth and he is well on his way of doing so. Consider what has been done to the vast riches of the seas.

With modern techniques, deep-sea fishing is so efficient that a few enterprises could rapidly sweep the oceans free of commercial fish. And this is what fishermen of all nationalities wish to do. As practical men they have no other interest than to use the latest technology that will increase their catch, preserve it and get it to market speedily as possible.

We witness at the moment the end of one of the saddest cases of misuse of technology by greedy fishing interests. Unless these interests are curbed by truly effective international action, the great whales—the blue, the finback, the sperm—will soon disappear, victims of man's "practical" folly.

These and other whales once populated the high seas in immense numbers. For hundreds of years whaling remained a reasonably fair contest between man and the intelligent, swift-moving mammals he hunted. Modern technology has turned it into brutal genocide. Blindly pursuing what they doubtless consider

an eminently *practical* objective—maximum profit *today*—the whalers are wiping out the very resources that could insure them a profit *tomorrow*.

In April of this year Japanese ships had to return home after only 3 of the normal 5 months at sea because they could find no whales.

Practical considerations aside, is anyone justified in using technology to exterminate a species that has existed on this earth for eons—the largest animal the world has ever seen? Are we certain our descendents may not at some future time have need of these mammals?

How we use technology profoundly affects the shape of our society. In the brief span of time—a century or so—that we have had a science-based technology, what use have we made of it? We have multiplied inordinately, wasted irreplaceable fuels and minerals and perpetrated incalculable and irreversible ecological damage. On the strength of our knowledge of nature, we have set ourselves above nature. We presume to change the natural environment for *all* the living creatures on this earth. Do we, who are transients on this earth and not overly wise, really believe we have the right to upset the order of nature, an order established by a power higher than man?

These are complicated matters for ordinary citizens to evaluate and decide. How in future to make wiser use of technology is perhaps the paramount public issue facing the electorates of industrial countries. It will tax their mental resources and challenge their political acumen. Certain measures suggest themselves:

Experience shows that by itself, the legal maxim of “the mutuality of liberty” will not prevent commitment to technologies that may later prove harmful. The maxim must be implemented by preventive public action—action of the kind that has long been operative in the field of public health. There is need for laws requiring that *before* a particular technology may be used, reliable tests must have been made to prove it will be useful *and* safe. A few such laws have been enacted; more are needed.

I suggest that, as a special public service, lawyers take on the task of working for better protection against technological injury. This is a new and fruitful area in which they could make important contributions to human welfare—an area which requires no revolutionary change in the political or economic structure of society, merely greater precision and fuller implementation of the traditional principle that injuring the health or causing the death of human beings is unlawful. The term “health” should not be limited to physical health but should include psychic health and protection of the human personality as well. New technologies based on the uncertain “science” of the social sciences involve snooping into the inner recesses of the human mind, personality testing and pseudo-scientific manipulation of human beings. When they are imposed as conditions of employment or otherwise partake of an element of compulsion, these technologies should be regulated or outlawed entirely.

Much more thought should be given to technological interference with the balance of nature and its consequences for man, present and future. There is need of wider recognition that government has as much a duty to protect the land, the air, the water, the natural environment against technological damage, as it has to protect the country against foreign enemies and the individual against criminals. Conversely, that every citizen is duty bound to make an effort to understand how technology operates, what are its possibilities, its limitations, its potential dangers. The leisure modern technology makes available to ever larger numbers of citizens could not be better spent than in a determined effort to narrow the knowledge gap between those who understand science and technology and those who do not.

Since law and public opinion always lag behind the swift development of new technologies, there is need for more informed and responsible thinking among those who control technologies. This might be achieved by professionalizing the decisionmaking process in technology. Experience has shown that in the hands of professional persons technology is managed with greater concern for human welfare than when it is controlled, as at present, by nonprofessionals. The classic example is medicine.

Of all technologies, that of the physician has benefited human beings most and harmed them least. The stringent standards set by the profession and by society for the education and professional conduct of physicians accounts for this happy circumstance. Not only is no one permitted to practice who has not given proof of his competence, but physicians must also be broadly, liberally, humanistically educated men and women. This gives them perspective in evalu-

ating their professional actions, an ability to see these actions against a humanistic background. Moreover, they operate under a code of ethics which requires them to place the needs of patients above all other considerations—a code incorporated 25 centuries ago in the Oath of Hippocrates, an oath still taken by young men and women embarking on a medical career.

To Greece we owe the noble idea that special knowledge and skill ought to be used to benefit man, rather than for personal aggrandizement or power, or as a means of extracting maximum gain from those in need of the services of men possessing special expertise. This concept of a *trusteeship of knowledge* could well be applied to *all* whose knowledge of science and technology surpasses that of the lay public, as it now is to physicians and surgeons. I have long advocated that engineering pattern itself after medicine and law, thus becoming a truly “learned” profession. It has, I believe, attained that status in some countries, though not in mine.

These are my suggestions; others may have better ones to offer. What seems to me of utmost importance is that we never for a moment forget that a *free society centers on man*. It gives paramount consideration to human rights, interests and needs. Society ceases to be free if a pattern of life develops where technology, not man, becomes central to its purpose. We must not permit this to happen lest the human liberties for which mankind has fought, at so great a cost of effort and sacrifice, will be extinguished.

#### QUESTIONS OF INVASION OF PRIVACY RELATING TO THE ESTABLISHMENT OF A NATIONAL DATA CENTER

[Reprinted from the Congressional Record, Aug. 18, 1966]

Mr. GALLAGHER. Mr. Speaker, under the direction of the Honorable William L. Dawson, chairman of the House Government Operations Committee, the Special Subcommittee on Invasion of Privacy, made up of my distinguished colleagues Congressmen Benjamin Rosenthal, of New York, and Frank Horton, of New York, has just conducted a series of hearings on the proposed establishment of a National Data Center by the Federal Government.

As chairman of this special subcommittee, it is my desire to present to the membership today some of my thoughts and findings upon the conclusion of those hearings.

Testimony before the subcommittee has illustrated first of all the great technological progress which has been made in the field of computer science and data processing. The potential of this technology and its value to our modern society are certainly impressive.

But the hearings have indicated as well an urgent need for a corollary study to determine the direction which our Nation will permit technology to take and the great responsibility we face to protect the public interest and rights of the individual.

Modern scientific achievement goes far beyond the full comprehension and knowledge of most of us. Yet its influence upon the life of each citizen is ever-increasing. As we realize our own inadequacy to evaluate an issue involving scientific technology, we seem to move toward an ever-increasing submission to the domination of those who are expert in the scientific disciplines.

Society borders on forgetting that technology is its own creation, to be guided and directed along the course which will provide its members most with the full benefits of scientific knowledge. The people seem dangerously prepared to surrender their age-old respect for the vast capabilities of the human mind and personality to the impressive and sometimes overwhelming knowledge which the scientific elite alone have mastered. Somewhat intimidated by the mystery of science, the average citizen in our Nation often seems reconciled to the sacrifice of individual liberties in the awesome name of “progress.”

Moreover, this malaise threatens to distort our traditional concept of law and its meaning to society. The forefathers of this Nation defined a legal code to protect the rights of Americans against government encroachment. The ultimate value of this code, which we know as our Constitution, centers in its flexibility, its adaptability to the needs of each new generation.

In science, however, the term “law” takes on new meaning. It defines the regularity of physical phenomena and its definitions seem synonymous with cold

fact. Scientific law is not concerned with the multiple aspects of free will and individual personalities; it deals solely with nature's constancy, or if you will—mechanical regularity.

In today's world, however, these two totally distinct definitions of one term seem to have fused and been confused. And in this commingling of definitions, it is indeed unfortunate that the scientific appears to have taken precedence.

This precedence is well illustrated in the case before us at present. The individual's right to privacy has always been recognized in the full course of our Nation's history. The protection of the individual by law against infringements attempted upon this right by others is incorporated in our Nation's Constitution. And yet in the name of scientific advancement, this right is now potentially threatened.

Although the proposal to establish a national statistical data center, if adopted, promises greater efficiency in many Government operations, the possibility that such a center might become a depository for extensive personal information on every citizen raises questions fraught with serious implications.

Should the Government establish a centralized statistical data center with its interchangeable counterpart a personal dossier bank, there would be a tremendous store of data already available to feed it. That list includes tax returns, census responses, social security data, military records, security files, fingerprints, FHA and VA mortgage guarantees, credit records, health data, and research involving individuals. If State and local governments were tied into the proposal, such data as school records, police files, driving violations, and property holdings would also be on file.

The Bureau of the Budget contends that no one has proposed such a personal dossier bank. That is quite true. But it is also a fact that detailed information on millions of individuals and corporations would be poured into the national data center. Group data is made up of individual data. Testimony by computer experts before our subcommittee shows clearly that a data center could easily become a dossier bank. Simply stated, our concern is what an innocent statistical center could turn into as the years roll by and pressure mounts to program into the computers more and more information on individuals. Computer experts stressed that the same technology that put the information in for statistical data could be retrieved instantaneously on any individual.

At present, the confidentiality of some of this information is protected by the law. Centralization, however, would create the need for a new set of safeguards to protect the privacy of the material on file. It would appear obvious that the Federal official who has the authority to press the button to produce a dossier on any individual in the United States would possess a power greater than any ever before known in America.

We must remember that our citizens give the Government personal information on a confidential basis and for a specific purpose. Americans deserve the assurance that this information will not be used for any other purpose in the future. Our Government must decide now before we embark on this new and dangerous course whether we can properly protect the civil rights and civil liberties of each citizen.

Without carefully established safeguards, there exists a very real threat of great injustice. Safeguards, which incidentally, do not now exist in a technological sense. It is certainly conceivable that a potential Big Brother—in the frightening Orwellian tradition—might make excellent use of a big button on a dossier bank for his own purposes and for the sake of increasing his own power.

Writing on this subject, the Wall Street Journal, August 5, 1966, stated:

"We do not suggest that many officials would attempt to abuse the power. Yet the fact is that even as it is, Federal agencies have been known to harass individuals or businesses, just as some of them have not been above electronic prying and other violations of privacy.

"It is a cardinal requirement of a free society that the people do not entrust their liberties to the whims of men in power but rely rather on wise laws to protect them from oppression."

It seems evident that if the proposal to create a national data bank is adopted, we will have to rely only on the hope that benevolent people with benevolent purposes will operate the system. History, however, has already taught a terrible lesson illustrating exactly what can happen when large stores of information become available to nonbenevolent powerseekers.

The detailed European census, long in effect even before the advent of the Nazi Party, provided a most convenient and efficient tool for Hitler's use when he led

the party to control Germany. The census information provided a central data system from which the dictator could draw detailed information on any German citizen, thereby facilitating the power surge of his totalitarian regime.

Prof. Charles A. Reich, Yale University expert on constitutional law, stated in his testimony before the Subcommittee on the Invasion of Privacy:

"Real protection in this world comes not from people's good intentions, but from the law."

The risk involved now in entrusting the liberties of the American people to the men of power in the future, the names of whom we do not even know and whose benevolence we cannot presume to guarantee, is too great for us to take.

We must consider now whether our laws are keeping abreast of our rapidly expanding technology and whether with the advance of that technology, our present safeguards of the constitutional rights of our citizens are still adequate.

In his address before the Royal National Foundation at Athens, June 1966, Vice Adm. H. G. Rickover, stated that in areas of technology, man must consider first whether any proposed action can be allowed legally and secondly, whether the action will be beneficial to his society. Admiral Rickover continued:

"These are matters that lie outside the domain of science. Just as the law of the cosmos cannot be overturned by human fiat, so is human law supreme within its own proper sphere of operation. Technology must therefore conform to that most basic of all human laws, the maximum of the 'mutuality of liberty,' the principle that one man's liberty of action ends where it would injure another."

In our modern society, therefore, we must meet the challenge of providing legal safeguards today against the possible violation by technology tomorrow of the maxim of the "mutuality of liberty." The supremacy of human law has been seriously questioned; it must be reasserted now if our society is to remain free.

The technology of computerization has raised new horizons of progress, but it also brings with it grave dangers. The greatest of these is that we may allow ourselves to drift into a course of action that will ultimately substitute a computer for man's free will and his human consciousness of what is ethical and what is not.

The 19th century novelist, Samuel Butler recognized the fallacy of such a course when he wrote in his novel, "Erewhon":

"I cannot think it will ever be safe to repose such trust in the moral sense of any machine."

Americans of the 20th century cannot think it any safer. As Professor Neisser has written in a study entitled "The Imitation of Man by Machine":

"If machines really thought as men do, there would be no more reason to fear them than to fear men. But computer intelligence is indeed 'inhuman': it does not grow, has no emotional basis, and is shallowly motivated. These defects do not matter in technical applications, where the criteria of successful problem-solving are relatively simple. They become extremely important if the computer is used to make social decisions, for there our criteria of adequacy are as subtle and as multiply as human thinking itself."

The temptation to utilize computer technology in more and more areas of public concern is great, of course, because it provides for a greater efficiency and accuracy in statistical studies and for a sense of scientific authority in solving difficult problems involving complex information.

However, in their article in a 1962 issue of Science magazine, Professors Johnson and Kohler warned against turning over to computers questions with which man is better able to cope:

"It (computer technology) is being called on to act for man in areas where man cannot define his own ability to perform and where he feels uneasy about his own performance—where he would like a neat, well-structured solution and feels that in adopting the machine's partial solution he is closer to the 'right' than he is in using his own."

There is certainly no doubt that computerized centralization of information gathered from the files of Federal agencies would facilitate many government operations. In the words of Johnson and Kohler, it would provide a "neat, well-structured solution" to the present unordered and amorphous task of maintaining complete and current files on information gathered from widely scattered sources for widely scattered purposes.

At the same time, however, in their article on "Privacy and Behavioral Research" that appeared last fall in Columbia Law Review, Oscar M. Ruebhausen and Orville G. Brim, Jr., noted:



"Computerized central storage of information would remove what surely has been one of the strongest allies of the claim to privacy—the inefficiency of man and the infallibility of his memory."

Thus, greater efficiency in Government operations would be paid for at the far greater expense of weakening the right to privacy of all American citizens. Surely this is too exorbitant a price to pay for an economized filing system.

In his essay entitled "Machinemade Justice: Some Implications," Prof. Joseph J. Spengler discussed certain implications of the use of computers in the definition and the administration of justice. His essay illustrated how and why certain biases may be introduced along with the use of computers in the area of law.

His final warning seems to the point:

"Great care must be exercised to avert the biases, the distortions, and the probable miscarriages of justice . . . If computers and other instruments are to be used, every precaution must be taken lest the mechanical servant become master, and a tyrannical one at that."

We are particularly concerned with the threat which a possible future Federal dossier bank represents to our Nation's basic Judeo-Christian doctrine which provides man with the unlimited right to make amends for his mistakes and to begin again. Man can forgive and forget the trespasses of his fellow man, but computers do not forget and they are incapable of forgiving. Rash actions one may have taken in one's youth and which one regrets in his maturity would be recorded for posterity on the computer's tape. Mistakes in judgment which are later regretted and repaired in an individual's personal and business life would remain imprinted indelibly on his computer data card. Thus, computerized files could become a bank of static, petrified and tyrannical information which can be used again and again to the disadvantage and harm of the American citizen.

We are now on the brink of making a fundamental change in our society which will destroy the basic philosophy of letting a man start anew, his record unblemished by past mistakes for which he has paid his just debt to society. We appear to be moving slowly but steadily toward a doctrine of complete scientific objectivity which will categorize and catalog each aspect of individuality, leaving as an end result a stack of computer program cards where once were human beings.

During the last decade, tremendous advancements have been made in the fields of psychology, sociology, political science and economics. Emerging under the new title of the "behavioral sciences," these disciplines have delved deeply into the complex problem of how man affects his society and how social institutions affect man. It is predicted that within 30 years, the behavioral scientists will be able to produce the achievers in our society at will. The August 15 issue of Newsweek reported that current research indicates that—

"The achieving child will be the product of order, home and school environment, and other factors under the control of parents—or the state."

Newsweek also predicted that within this time period, sociologists will have developed the "complete picture of manpower flow in our Nation." Orville Brim, president of the Russell Sage Foundation—a private institution for research—was quoted as saying:

"Incentives like money, and educational opportunity could be controlled so that people are properly distributed."

The prospect of such social manipulation is a fascinating one, but it also raises serious questions, the answers to which are at present unknown. Who will define a "proper distribution" of financial and educational opportunity and who can guarantee that society will use such manipulative techniques constructively?

Experts in the field of computer science have joined with behavioral scientists, combining the knowledge of both disciplines in conducting extensive research into the question of man's relationship to the society in which he lives. The widespread use of computers has greatly facilitated the accumulation of data and the transformation of varied and unrelated information into clear and meaningful statements on man's behavior and the pattern of his actions in the past.

But it is essential to remember that the computer cannot predict infallibly what a man will do in the future; it cannot set down axiomatic laws to govern a man's action at any given moment in the next hour, the next day or the next year. A computer cannot measure courage, loyalty or love.

More than 2,000 years ago Plato wrote:

"The differences of men and actions, and the endless irregular movements of human beings, do not admit of any universal and simple rules. And no art whatsoever can lay down a rule which lasts for all time."

It is our greatest fear that modern computer technology will attempt to do just that—to establish on the basis of compiled data on man's past actions axiomatic principles for predicting what he will do in the future—and that these principles will become accepted by society as nearly infallible. The final result would be the restriction of a man's future based upon the statistical pattern of his actions in his youth.

I believe that our Government and our legal structure must act now to retain their flexibility, if they are to remain useful and beneficial to society. Both must continue to recognize and respect the infinity of choices which a single man may make during the course of a single life-span. This is why we voice so strongly our opposition to the present proposal to establish a national data bank which would store for all time private and confidential information on all citizens which might later be used to restrict their free choice of action in the future.

Therefore, it is essential that we begin to determine now the potential of today's technology and how our traditional liberties and beliefs can be protected from a technological onslaught in the future. Most specifically, we must chart today the course we want that will allow computer science to follow in contemporary society, and in our society of tomorrow.

We must call upon the scientific community, which is responsible for the development of this technology, to bear an equal responsibility for its control, in order to guarantee adequate protection of the freedoms we now enjoy. For this reason, it is of vital importance that representatives from all of the disciplines involved in the development and implementation of the computer join in creating a symposium independent of any Federal agency to examine the potential of computer science and its effects on the rights of the individual.

The need for such a symposium was well indicated during the course of our hearings last month. Spokesmen for the Bureau of the Budget came to the subcommittee to discuss the establishment of a Statistical Data Center. Yet, under extensive interrogation, the witnesses proved to be at a distressing loss of words when pressed for a detailed explanation of the system and for specific safeguards that could be built into the center to provide for the protection of the individual's right to privacy.

They seemed unable to comprehend the ease with which a statistical data bank could be converted into a personal dossier center, and they failed to realize the potential power of such a center. This is, we believe, the crux of the problem with which we are faced. There appears to be a basic lack of communication among the computer scientists, the behavioral scientists and experts in constitutional law and civil liberties.

It is unfortunate that while the American Orthopsychiatric Association was discussing data banks and invasion of privacy at its annual convention in San Francisco in April, the American Statistical Association was planning its convention for later this month in Los Angeles with a panel discussion on the "design and use of statistical data banks." The New York City Bar Association's Committee on Science and Law was studying the "impact of science and technology on privacy," and the American Bar Association was planning a special section of the association to deal with legal problems concerning individual rights and particularly the relationship of these rights to modern science. Yet, apparently no effort has been made on the part of these individual associations to combine their interests in a joint symposium.

We can no longer afford isolated contemplation in this area. Certainly social scientists, computer technologists and experts in constitutional law are all equally concerned with the problem of achieving a balance between advancing technology and the preservation of individual liberties. The computer has made all of us partners in the development of modern technology and we must continually educate each other, if we are to achieve this crucial balance. A symposium, conducted by experts from each of these fields could serve as an arena for discussion of all major aspects of the problem.

It is of utmost importance that the symposium consider, independently of the Federal Government, the technical, legal, and sociological aspects of the proposal which the Government seems intent upon rationalizing.

The symposium might be conducted through a series of seminars organized by professional associations such as the American Economic Association, the



Association for Computing Machinery or the American Bar Association. Sponsorship of such a symposium might also be offered by a consortium of American universities, for certainly the role of the scholar as independent social commentator has long been traditional in our universities and colleges. Those private foundations and institutions which have financed research in these areas during the past decade might be called upon to fund the symposium.

In seminar fashion, the symposium would consist of a series of papers submitted and discussed by panels comprised of experts from each of the disciplines involved. It is hoped that the proceedings of the seminars could be published and widely circulated in order to provide for even more extensive examination by other members of the computer industry, the academic world, and the public in general.

The findings of such a symposium would then represent the recommendations formulated by the most qualified experts in our Nation. Only with the benefit of their accumulated knowledge could Congress properly evaluate such proposals as the Bureau of the Budget's, calling for the creation of a National Data Center. Only when all of the implications raised by such a proposal have been examined objectively can Congress act effectively to insure its legality.

In this era which has been so greatly affected by the machine, we must call upon our greatest natural resources—that is, the wisdom and knowledge of America's scientists, scholars and legal experts—to aid us more than ever before in achieving a balance between the advancement of scientific technology and the preservation of constitutional liberties.

The New York Times of August 9 in commenting on our hearings and expressing their concern over the National Data Center summed it all up in two lines:

"Perhaps in the long run the fight to preserve privacy is a vain one. But, like the struggle to preserve life it must be continued while any shred of privacy remains."

In the future no responsibility of our Government will be greater than the preservation of privacy and the protection of our fundamental human values. Like the problem of nuclear warfare, it is a time to reflect on how far we have come before we drift into a course that is beyond our capacity to navigate.

Mr. HORTON. It is a privilege for me to join the gentleman from New Jersey [Mr. GALLAGHER], chairman of our Special Subcommittee on the Invasion of Privacy, in pointing out some of the findings of our recent hearings on the proposed National Data Center. Chairman GALLAGHER has competently reviewed the fears and feelings these hearings have fostered: Scientific technology is the creation of man, and he, as creator, must also be controller. Before we allow ourselves to be mastered by machines, we must consider our rights guaranteed by the Constitution and take the necessary action to protect them.

Though a dossier bank is not proposed, testimony in our hearings substantiates the supposition that a statistical center, which is proposed, does have the potential to hold privacy-invading information on the citizens of this country. The spectrum of information already contained in computers within the separate agencies, if brought together, could reveal with the push of one button every record made on an individual from his birth certificate to the present. In fact, there is no reason to doubt that such a system could trace back and even bring together information on an individual's parents, grandparents, aunts, uncles, cousins, friends, and associates.

As I pointed out in my opening statement at our hearings, a central data service bank would require:

First. That confidential information now in Government files would be forwarded to a new group and used for other purposes than it was originally given; and

Second. That a new group would have the code and would have access to the names, addresses, and background of the people to whom this confidential information relates.

Tying the two together would be an easy matter—and it would be an outright denial of our right to privacy. As Vance Packard brought to our attention during the hearings, when the social security number was originated, it was a confidential reference. Now it is requested and given on practically every form an individual completes in his lifetime. It seems to grow easier to give out information, whether or not it was once confidential, than it is to protect confidentiality. One of the greatest safeguards now protecting information possessed by various agencies is its fragmented nature. Retrieval is impractical and often impossible. A central data bank removes completely this

safeguard and threatens to make available entire dossiers, rather than simply a social security number.

An injustice which deepens the privacy invasion constituted by such an information center is the individual's complete ignorance of what is contained in his dossier—whether it is scattered throughout various agencies as it is now, or available in one bank. Certainly it is not unreasonable that individuals be given the opportunity to see their files and to rebut information they feel is fallacious or as it may often be, the result of prejudiced reports.

On July 28 I issued a statement supporting the idea for a symposium of educators, social scientists, lawyers and computer technologists to study methods to prevent large computerized data banks from invading the privacy of the American citizen. I would like to share with my colleagues the text of my remarks regarding this Government-sponsored symposium:

"Computer technology is advancing so quickly that we are in danger of allowing machines to bare the most private facts in our lives.

Our hearings have pointed out a tremendous need for knowledge in this area of computers and privacy. This symposium would fill that gap and allow us to work out adequate safeguards against unauthorized release of private information to people with no right to have it.

Vance Packard, a witness at our hearings, warned that 'Big Brother, if he ever comes to the United States, may turn out to be not a greedy power-seeker but rather a relentless bureaucrat obsessed with efficiency. And he, more than the simple power seeker, could lead us to that ultimate of horrors, a humanity in chains of plastic tape.'

Packard, whose best-selling books have documented some of the country's social ills, said, 'The filekeepers of Washington have derogatory information of one sort or another on literally millions of people. The more such files are fed into the central files, the greater the hazard the information will become enormously tempting to use as a form of control.'

Computer technology is so complex that we can't pass adequate laws immediately to block release of private information. This symposium must study the question with the aim of defining the problems involved.

This committee should be commended, since it is one of the first groups to recognize the dangers inherent in the use of the giant, information-storage devices. For, as another witness at our hearings, Paul Baran of the Rand Corporation, predicted, computers of the foreseeable future could be 10,000 times the size of those available today.

As I said during the hearings, there is a danger that computers, because they are machines, will treat us as machines. They can supply the facts and, in effect, direct us from birth to death. They can 'pigeonhole' us as their tapes decree, selecting within a narrow range the schooling we get, the jobs we hold, the money we earn and even the person we marry.

It is not enough to say 'it can't happen here'; our grandfathers said that about television."

It is clear that a computer bank offers great potential for advancements in efficiency; yet, it also offers great potential for invasion of privacy. I do not propose to limit the progress of science or knowledge, but to control the use of knowledge scientific advancement makes available so readily. The concept of such control is not new; man's greatest helpers can also be his greatest destroyers if not controlled—we need only consider basic elements as water and fire to confirm this.

Mr. Speaker, I want to conclude by inserting an editorial entitled "To Preserve Privacy," which appeared in the August 9 edition of the New York Times. Following the editorial is the text of a letter I wrote to the editor commending his editorial and reviewing the necessity to protect the freedom of individuals with private lives:

#### "TO PRESERVE PRIVACY"

"Can personal privacy survive the ceaseless advances of the technological juggernaut? Many in public and private life now fear to use telephones for conversations they would keep confidential, while the variety of electronic 'bugs' available to eavesdrop on even whispered communications staggers the imagination. And young lovers would be well-advised to remember that the skies are increasingly full of sputniks equipped with cameras capable of taking extraordinarily detailed pictures of what transpires under the moon as well as on it. George Orwell foresaw the logical end of this trend in a device that would enable 'Big Brother' to keep an eye on everyone anywhere.

"The Orwellian nightmare would be brought very close indeed if Congress permits the proposed computer National Data Center to come into being. We already live with the fact that from birth to grave Federal agencies keep tabs on each of us, recording our individual puny existence, monitoring our incomes and claimed deductions, noting when we are employed or jobless, and—through the F.B.I. and similar agencies—keeping all too close watch on what we think or say, what we read and what organizations we belong to.

"If this situation is still somewhat tolerable, it is because each agency keeps separate files and it takes some considerable effort to find and bring together all that is known about a particular individual. What is now proposed is the amalgamation of these files, and the creation of a situation in which the push of a button would promptly dredge up all that is known about anyone.

"Understandably, this idea has brought vigorous protest, in which we join. Aside from the opportunities for blackmail and from the likelihood that the record of any single past transgression might damage one for life, this proposed device would approach the effective end of privacy. Those Government officials who insist that the all-knowing computer could be provided with safeguards against unauthorized access are no doubt of the same breed as their brethren who 'guaranteed' that last November's Northeast electric blackout could never occur. Even the Swiss banks have learned to their own and their clients' sorrow that the device of numbered accounts is inadequate to frustrate determined would-be blackmailers.

"Perhaps in the long run the fight to preserve privacy is a vain one. But, like the struggle to preserve life, it must be continued while any shred of privacy remains."

AUGUST 10, 1966.

LETTERS TO THE EDITOR,  
*The New York Times*,  
New York, N.Y.

DEAR SIR: As a citizen and Congressman, and particularly because of my service as the ranking and only minority member of the Special Subcommittee on Invasion of Privacy of the House Government Operations Committee, I commend your August 9 editorial, "To Preserve Privacy."

Our Subcommittee, chaired by Congressman Cornelius Gallagher of New Jersey, just completed hearings on the Federal Government proposal to establish a National Data Center. The testimony removed any doubt as to the foundation of your "1984" fear.

However, I am not willing to resign myself to the fateful suggestion that "Perhaps in the long run the fight to preserve privacy is a vain one." Rather, I intend to continue, as your editorial also challenges Congress, with my struggle for its protection.

The problems posed by the National Data Center proposal are in carload lots. Like nuclear energy, there is nip and tuck competition to keep computers working for us, not against us.

There is undeniable value from the standpoints of economy and efficiency in allowing agencies to pool data. But, if it is to be done, must there not be ground rules and clear-cut standards?

Thus, I recommend the calling of a symposium of computer men, sociologists, educators, lawyers, and others to identify a correct course. These men and women should suggest safeguards: coding, surprise audits, inter-computer interrogation limits, interrogator identity, abnormal interrogation detection, illegal disclosure penalties, and many more. We also need to consider the individual's right to know the contents of any government dossier on him.

Computers can give us longer life, teach our students, design better transportation, provide our statesmen with facts and figures for sound decisions, capture criminals, diagnose disease, and add new dimensions to every element of society. But, making it possible means maintaining man as the master of the machine.

Advanced technology must not be paid from the accounts of individuality. For as Justice Brandeis said, we all are entitled "to be let alone."

Sincerely,

FRANK HORTON,  
U.S. Congressman, 36th District of New York.